UNIVERSIDADE DO SUL DE SANTA CATARINA

CHRISTIAN MUNIZ DO VALLE PENHA

LUCAS MONIR HENRIQUE

PERSPECTIVAS FUTURAS DA INTELIGÊNCIA ARTIFICIAL E O APRENDIZADO DE MÁQUINA: DISCUSSÕES SOBRE O USO DA INTELIGÊNCIA ARTIFICIAL GENERATIVA

Florianópolis

2023

CHRISTIAN MUNIZ DO VALLE PENHA

LUCAS MONIR HENRIQUE

PERSPECTIVAS FUTURAS DA INTELIGÊNCIA ARTIFICIAL E O APRENDIZADO DE MÁQUINA: DISCUSSÕES SOBRE O USO DA INTELIGÊNCIA ARTIFICIAL GENERATIVA

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Sistemas de Informação da Universidade do Sul de Santa Catarina, como requisito parcial à obtenção do título de Bacharel em Sistemas de Informação.

Orientador: Prof. Saulo Popov Zambiasi, Dr.

Florianópolis

2023

CHRISTIAN MUNIZ DO VALLE PENHA

LUCAS MONIR HENRIQUE

PERSPECTIVAS FUTURAS DA INTELIGÊNCIA ARTIFICIAL E O APRENDIZADO DE MÁQUINA: DISCUSSÕES SOBRE O USO DA INTELIGÊNCIA ARTIFICIAL GENERATIVA

Este Trabalho de Conclusão de Curso foi julgado adequado à obtenção do título de Bacharel em Sistemas de Informação e aprovado em sua forma final pelo Curso de Graduação em Sistemas de Informação da Universidade do Sul de Santa Catarina.

Florianópolis, 14 de junho de 2023.

_____

Prof. Saulo Popov Zambiasi, Dr.
Universidade do Sul de Santa Catarina

_____

Prof. Nome do Professor, abreviatura da titulação.
Universidade do Sul de Santa Catarina

_____

Patricia Leandra Barrufi Pinheiro, Drª.

Dedicamos este TCC aos nossos pais, amigos e a todos que nos apoiaram em nossa jornada acadêmica.

AGRADECIMENTOS

Agradecemos a nossos amigos e familiares pelo apoio e incentivo em nossa jornada acadêmica e profissional, e também ao nosso professor e orientador Saulo Popov Zambiasi por sempre estimular nossa criatividade e interesses nas aulas.

"*Podemos ver pouco à frente no futuro, mas podemos ver que há muito que precisa ser feito*" (TURING, 1950).

RESUMO

O crescimento rápido da utilização da inteligência artificial no dia a dia traz impactos diretos e indiretos a sociedade. Plágio, desinformação e difamação são alguns exemplos de problemas que surgem com o paradigma de tecnologias próximas de autossuficientes. Assim sendo, foi feita a contextualização das tecnologias e técnicas de criação de mídia sintética, além da agregação e reorganização de notícias, dados e informações sobre como mídia sintética gerada por modelos de inteligência artificial podem causar danos à sociedade se usadas para más intenções. Com o conhecimento obtido pela pesquisa, uma nova perspectiva foi obtida onde o avanço tecnológico não deve sobressair o bem-estar social e a utilização de informações para treinamento de IAs deve ser curada não só para fornecer bases de melhor qualidade mas também para respeitar os direitos dos artistas que criaram a arte.

Palavras-chave: Inteligência artificial. Mídia sintética. Discussão crítica.

ABSTRACT

The rapid growth in the use of artificial intelligence in everyday life brings direct and indirect impacts to society. Plagiarism, misinformation, and defamation are some examples of problems that arise with the paradigm of technologies approaching self-sufficiency. Therefore, a contextualization of synthetic media technologies and creation techniques was made, along with the aggregation and reorganization of news, data, and information on how synthetic media generated by artificial intelligence models can cause harm to society if used with malicious intent. With the knowledge obtained from the research, a new perspective was gained where technological advancement should not outweigh social well-being, and the use of information for AI training should be carefully handled not only to provide higher quality foundations but also to respect the rights of the artists who created the art.

Keywords: Artificial intelligence. Synthetic media. Critical discussion.

LISTA DE ILUSTRAÇÕES

FIGURA 1 - Representação da IA e suas camadas como bonecas “Matryoshka” **21**

FIGURA 2 - Interface, prompt de texto e imagens geradas a partir de texto com DALL-E **26**

FIGURA 3 - Interface de usuário do Dall-E com botões para manipulação da imagem **27**

FIGURA 4 - Imagem criada usando o DALL-E com o prompt “cão e pássaro tomando chá ao estilo surrealista em watercolor” **28**

Figura 5 - Imagem criada usando o DALL-E com o prompt “dois ursos de pelúcia fechando negócio em uma cidade cyberpunk iluminada por neon.” **29**

Figura 6 - Imagens geradas via demo online do Stable Diffusion, com o prompt “coelho pulando sobre um arco-íris” **30**

Figura 7 - Imagens geradas via demo online do Stable Diffusion, com o prompt “Johnny Cash jogado Futebol” **31**

Figura 8 - “Magic Avatar” da redatora Zoe Sottlie, CNN, utilizando o Lensa **32**

Figura 9 - ChatGPT respondendo uma pergunta sobre “O que é ChatGPT?” **33**

Figura 10 - ChatGPT avisando das possibilidades de informações incorretas, com viés ou enganosas **34**

Figura 11 - ChatGPT descrevendo o aquecimento global como uma música de Raul **Seixas 35**

Figura 12- ChatGPT descrevendo a si mesmo no estilo de Edgar Allan Poe **36**

Figura 13 - Imagem gerada pelo Midjourney para o prompt “uma casa feita de cogumelos amanita muscaria realista” **37**

Figura 14 - IA Midjourney disponibilizada via Discord **37**

LISTA DE GRÁFICOS

GRÁFICO 1 - PARCELAS DAS ÁREAS DE ESTUDO ONDE I.A É IMPLEMENTADA NA EDUCAÇÃO **35**

LISTA DE TABELAS

TABELA 1 – Definições de IA separadas em quatro categorias **19**

TABELA 2 – Frequência da utilização de IA em sistemas ITS **35**

SUMÁRIO

1. INTRODUÇÃO....15
1.1. PROBLEMÁTICA....16
1.2. OBJETIVOS....17
1.2.1. Objetivo Geral....17
1.2.2. Objetivo Específico....17
1.3. JUSTIFICATIVA....18
1.4. ESTRUTURA DA MONOGRAFIA....18
2. FUNDAMENTAÇÃO TEÓRICA....20
2.1. A INTELIGÊNCIA ARTIFICIAL....20
2.2. AGENTES INTELIGENTES....22
2.3. MACHINE LEARNING....23
2.4. DEEP LEARNING....24
2.5. REDES NEURAIS....25
2.6. TECNOLOGIAS....25
2.6.1. Dall-E....25
2.6.2. Stable Diffusion....30
2.6.3. Lensa AI....32
2.6.4. Chat-GPT....33
2.6.5. Midjourney....36
2.6.6. Deepfake....38
2.7. APLICAÇÕES DA INTELIGÊNCIA ARTIFICIAL NA MEDICINA....39
2.8. APLICAÇÕES DA INTELIGÊNCIA ARTIFICIAL NA EDUCAÇÃO....40
2.9. APLICAÇÕES DA INTELIGÊNCIA ARTIFICIAL NA SEGURANÇA....42
3. MÉTODO....44
3.1. TIPO DE PESQUISA....44
3.2. ETAPAS UTILIZADAS....45
3.3. DELIMITAÇÕES....45
4. DISCUSSÕES SOBRE O USO DA INTELIGÊNCIA ARTIFICIAL GENERATIVA 47
5. CONCLUSÕES E TRABALHOS FUTUROS....55
5.1. TRABALHOS FUTUROS....55
REFERÊNCIAS....57

1. INTRODUÇÃO

O termo inteligência artificial é um leque muito abrangente. Para discutir sobre suas tecnologias, antes é necessário contextualizar o que é inteligência artificial e como ela pode ser interpretada por quem entra em contato com o termo. Para alguns, inteligência artificial é o ato de uma máquina agir de forma inteligente, enquanto para outros, é uma área da ciência. Sendo assim, será seguida a definição de McCarthy (2004), em que o mesmo afirma que a inteligência artificial:

> "[...]é a ciência e engenharia de fabricar máquinas inteligentes, especialmente programas de computador inteligentes. Ela está relacionada à tarefa semelhante de usar computadores para entender a inteligência humana, mas a IA não precisa se limitar aos métodos biologicamente observáveis." (MCCARTHY, 2004, p. 2).

Com o advento de métodos de aprendizado para inteligência artificial como *machine learning*, *deep learning* e *redes neurais*, criações antes vistas como exclusivamente humanas conseguem ser recriadas e derivadas, produzindo conversações, narrativas, imagens e diversas outras formas de conhecimento. Ainda assim, segundo Norvig e Russell (2013) este comportamento não é observado como autonomia ou consciência, sendo simplesmente informações criadas a partir do estudo de bilhões de exemplos alimentados ao sistema que os disseca, analisa e arquiva-os como exemplos a serem seguidos quando necessário, categorizando como racionalidade.

A API GPT-3 criada pelo laboratório de pesquisa em inteligência artificial OpenAI, consegue reconhecer, tratar e gerar informações e conhecimentos baseados em linguagem natural. Exemplos dados no site do laboratório incluem criação de slogan a partir de uma explicação com frases, criação de resumo de texto, formatação de texto, interpretação de texto e tradução, com todas essas funções sendo ensinadas pelo método *deep learning.* (OpenAI, 2022).

Esta tecnologia encaixa-se no termo geral "inteligência artificial generativa" que, como afirma a GenerativeAI (2023), é um tipo de IA capaz de gerar diversos tipos de dados, como imagens, vídeos, áudio, texto e modelos 3D. Várias IAs generativas notáveis, como o ChatGPT e o DALL·E são acessíveis gratuitamente online. Essas I.As possuem a capacidade

de criar conteúdo digital e são treinadas em grandes conjuntos de dados, aprendendo a capturar padrões e estruturas para produzir saídas criativas e coerentes.

O surgimento da internet e sua subsequente adição à rotina cotidiana da sociedade humana marcam o momento onde a presença de IAs em aplicações de grande escala conectadas à web se tornou norma devido a quantidade exorbitante de dados e interações necessárias para o melhor funcionamento do ecossistema on-line fazendo ressurgir o interesse pela área da ciência da computação. De *search engines*, *recommender systems* e *data miners* até criação de obras de arte baseadas em linguagem natural, a evolução da inteligência artificial tomou patamares anteriormente imaginados como exclusivos para mundos irreais de ficção científica e continua a se desenvolver ao ponto de pesquisadores responsáveis pelo seu desenvolvimento avisarem sobre seus riscos. (Brown, et al, 2020).

1.1. PROBLEMÁTICA

Mesmo com o avanço positivo que a inteligência artificial vem apresentando desde sua concepção, continuam sendo comuns as discussões sobre as incertezas de tal tecnologia, destacando-se os tópicos de segurança, ética e privacidade.

Tais discussões vêm ganhando força, como aponta a pesquisa feita por Liang, Y., Lee, S.A. (2017) cerca de 26% dos entrevistados têm medo elevado de robôs autônomos e inteligências artificiais. Em uma conclusão da pesquisa feita por Melanie Arntz (2017) em seu artigo “Revisiting the risk of automation”, é estimado que 9% dos empregos nos Estados Unidos corram risco de automação, tal medo pode ser justificado ainda mais quando é levado em consideração que as IAs, por terem ganho maior popularidade, e por conta do sistema legal da maioria dos países, possuem pouca regulamentação vigente em 2022, sendo a proposta precursora vinda da União Europeia em 2021, notável pela repercussão que causou ao visar regulamentar o uso de inteligência artificial no território europeu, impulsionando outros países a fazer o mesmo. Dentre os países que seguiram o exemplo europeu está o Brasil com o projeto de lei PL 21/2020, que institui diretrizes para o desenvolvimento e uso das tecnologias relacionadas à inteligência artificial no país.

Para garantir uma melhor adaptação à sociedade e ao mercado, é crucial desmistificar os perigos da inteligência artificial e, ao mesmo tempo, destacar os riscos reais decorrentes do avanço dessa tecnologia.

1.2. OBJETIVOS

Nessa seção são apresentados o Objetivo Geral e os Objetivos Específicos

1.2.1. **Objetivo Geral**

Este trabalho tem o objetivo de fazer uma explanação sobre o contexto da Inteligência Artificial Generativa na criação de conteúdo e levantar este tópico para uma discussão geral, tanto dos problemas envolvidos quanto para as perspectivas futuras dessa tecnologia.

1.2.2. **Objetivo Específico**

Os objetivos específicos deste projeto são:

- Obter referencial teórico e histórico sobre a inteligência artificial com foco principalmente na IA Generativa;
- Analisar o crescimento da IA no mercado;
- Discutir sobre os impactos da IA generativa.

1.3. JUSTIFICATIVA

À medida que as tarefas se tornam cada vez mais complexas e além das capacidades humanas, surge a necessidade de agentes racionais artificiais no contexto empresarial, especialmente em produtos voltados para consumidores comuns.

Entretanto, a utilização de IAs não se limita apenas ao tratamento de dados por interações de usuários, de acordo com a pesquisa feita pela empresa McKinsey (2022), “[organizações] que usam mecanismos de crescimento de vendas B2B orientados por dados reportam crescimento acima do mercado e aumento de lucro na faixa de 15% a 25%” .

Dados possuem imenso valor de mercado e têm o potencial de manter organizações centenárias firmes e fortes no mercado, acompanhando tendências e agilizando a produção.

1.4. ESTRUTURA DA MONOGRAFIA

Este trabalho está dividido em 5 capítulos:

O primeiro capítulo contém a introdução, descrição do problema, área de estudo, tema e objetivos da obra.

O segundo capítulo é dedicado à fundamentação teórica do trabalho, que foi feita por meio de artigos, notícias e obras literárias relacionadas à inteligência artificial.

O terceiro capítulo é dedicado a descrever a metodologia de pesquisa utilizada para levantamento de informações e para elaboração das conclusões presentes neste trabalho.

O quarto capítulo é focado em apresentar uma discussão crítica de mídias geradas por inteligência artificial, baseada nos dados levantados anteriormente no segundo capítulo.

O quinto e último capítulo apresenta as conclusões alcançadas durante a criação do trabalho e também possíveis trabalhos futuros.

2. FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta uma base teórica sobre a inteligência artificial como ciência e a conceptualização de agentes racionais e seus métodos de aprendizado, além também de apresentar tecnologias novas e já estabelecidas que utilizam estes métodos.

2.1. A INTELIGÊNCIA ARTIFICIAL

Segundo McCarthy (2004), inteligência é a parte computacional da habilidade de alcançar objetivos no mundo, ocorrendo em pessoas, animais e algumas máquinas. Sendo assim, a inteligência não é limitada ao contexto humano e tão menos ao contexto biológico.

Ainda de acordo com McCarthy (2004), a inteligência artificial não está isolada apenas a simulação do pensamento e inteligência humana, mas sim a resolução de problemas apresentados pelo mundo de forma inteligente. Sendo assim, pesquisadores são livres para utilizar quaisquer métodos para alcançar inteligência ou comportamento considerado inteligente ao resolver problemas apresentados pelo mundo, mesmo que estes utilizem habilidades computacionais que pessoas não conseguem ter.

Na visão de Russell e Norvig (2013), inteligência artificial é vista de forma multidimensional, com pesquisadores sendo a variante que determinam os métodos, objetivos, e finalidades. A tabela 1 mostra quatro das definições de pensamentos da inteligência artificial com exemplos.

Tabela 1 - Definições de IA separadas em quatro categorias

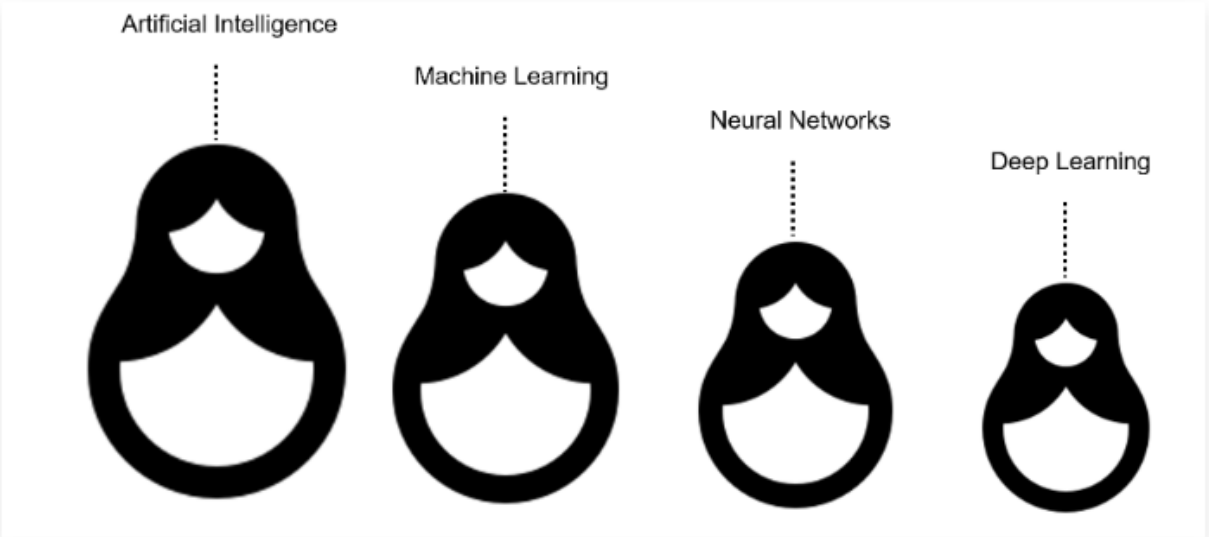
Pensando como um humano	**Pensando racionalmente**
“O novo e interessante esforço para fazer os computadores pensarem (...) máquinas com mentes, no sentido total e literal.” (Haugeland, 1985) “[Automatização de] atividades que associamos ao pensamento humano, atividades como a tomada de decisões, a resolução de problemas, o	“O estudo das faculdades mentais pelo uso de modelos computacionais.” (Charniak e McDermott, 1985) “O estudo das computações que tornam possível perceber, raciocinar e agir.” (Winston, 1992)

aprendizado...” (Bellman, 1978)	
Agindo como seres humanos	**Agindo racionalmente**
“A arte de criar máquinas que executam funções que exigem inteligência quando executadas por pessoas.” (Kurzweil, 1990) “O estudo de como os computadores podem fazer tarefas que hoje são melhor desempenhadas pelas pessoas.” (Rich and Knight, 1991)	“Inteligência Computacional é o estudo do projeto de agentes inteligentes.” (Poole et al., 1998) “AI... está relacionada a um desempenho inteligente de artefatos.” (Nilsson, 1998)

Fonte: Russell e Norvig (2013, p. 25)

Segundo IBM Cloud Education (2020), inteligência artificial nos tempos contemporâneos se destacam pelos seus métodos de aprendizado e como estão interligados, comparando os diversos métodos a bonecas russas “matryoshka”, como mostra a figura 1. Essa representação ajuda a explicar como novos descobrimentos na área levam a outros novos descobrimentos, em uma cadeia de “camadas”.

Figura 1 - Representação dos descobrimentos da IA e suas camadas como bonecas “Matryoshka”



Fonte: IBM Cloud Education (2020, online).

A figura 1 exemplifica como estes descobrimentos estão interligados, tendo técnicas, métodos e novas formas de aprendizado sendo descobertas dentro de outras, como bonecas “Matryoshka”, bonecas russas de tamanhos decrescentes as quais estão organizadas uma dentro da outra, da menor para a maior, semelhante a forma como redes neurais se

desenvolveram do aprendizado de máquina e como o *Deep Learning* surgiu com base nas redes neurais.

2.2. AGENTES INTELIGENTES

Segundo Norvig e Russell (2013), inteligência artificial é o estudo de agentes que tomam decisões baseados em percepções do ambiente. Estes agentes, referidos como agentes inteligentes ou racionais, são algo que age de uma forma.

Estes agentes que, segundo Norvig e Russell (2013), agem de forma racional com base em percepção do ambiente e seus arredores, seja através de sensores que interagem com o mundo físico ou camadas de entrada que interagem com o ambiente tecnológico, possuem racionalidade e serem definidos como inteligentes.

Ainda assim, Norvig e Russell (2013) apontam para o fato de que racionalidade e inteligência não estão atrelados aos conceitos humanos de algo inteligente. Racionalidade significa que algo ou alguém age de forma racional, ou seja, age para obter o melhor resultado de acordo com o que é esperado. Se um software é programado para lidar com linguagens naturais e o mesmo consegue interpretar, sintetizar e escolher a melhor resposta, o mesmo é considerado racional, mesmo que este não possua consciência ou noção de ser.

O teste de Turing, proposto em 1950 pelo cientista da computação Alan Turing, foi projetado com o intuito de dar uma resposta satisfatória à questão sobre a inteligência de computadores. Norvig e Russell (2013), abordam que o teste de Turing, que funciona ao possuir um investigador fazendo perguntas por escrito e então deduzindo se as respostas vieram de um computador ou ser humano, é um método relevante a história da inteligência artificial, mas que não devem ser o foco no desenvolvimento de novos agentes inteligentes. Ao possuir um agente inteligente limitando-se a imitar humanos, o programa se torna excessivamente complexo e específico para encaixar-se no teste.

Assim sendo, Norvig e Russell (2013) afirmam que agentes inteligentes devem ser tratados como protótipos de aviões no início da aviação. Substituindo a necessidade de imitar o orgânico por adaptar e utilizar dos pontos únicos da tecnologia, obtendo o resultado desejado por meios e métodos diferentes.

2.3. MACHINE LEARNING

De acordo com IBM Cloud Education (2020), o termo Machine Learning, conhecido em português como “Aprendizado de máquina", foi utilizado pela primeira vez pelo cientista da computação Arthur L. Samuel em sua pesquisa *“Some studies in machine learning using the game of checkers”*, de 1959, onde o termo se refere a como uma máquina é capaz de aprender com base nos dados fornecidos à ela.

Segundo IBM Cloud Education (2020), Machine Learning utiliza métodos de aprendizado através de algoritmos e dados para imitar a forma que seres humanos aprendem, aumentando sua precisão gradualmente. Esta precisão pode ser obtida através do “treinamento” da IA, o ato de suprir os dados que serão analisados e manipulados pelos algoritmos, criando um banco de dados dedicado à finalidade da aplicação.

De acordo com Shalian (2018), o aprendizado de máquina pode ser separado em quatro categorias, sendo elas:

- **Aprendizado supervisionado**: Os dados de controle são pré-categorizados e agrupados por usuários, fornecendo uma fonte confiável e correta para o treinamento do algoritmo.
- **Aprendizado não-supervisionado**: Os dados são fornecidos de forma “crua”, sem categorização ou agrupamento e então o algoritmo extrai padrões úteis ou comuns entre os dados de treinamento.
- **Aprendizado semi-supervisionado**: Os dados possuem exemplos tratados e não tratados, deixando o algoritmo tomar decisões independentes para categorizar os dados não tratados.
- **Aprendizado reforçado**: Utiliza um sistema semelhante a “recompensa e punição”. Os dados são fornecidos sem tratamento e o sistema é treinado a partir do feedback baseado em seus acertos e erros.

UC Berkeley (2020) quebra os algoritmos de aprendizado de máquina supervisionados em três etapas, sendo estas: um processo de decisão, uma função de erro e um processo de otimização ou atualização.

Um processo de decisão é um conjunto de cálculos ou outros passos que manipulam os dados e tentam "adivinhar" qual padrão os algoritmos tentam encontrar. A função de erro é responsável por medir o quão boa foi a tentativa de adivinhar, comparando

com exemplos corretos. Caso esteja errado, apresenta uma forma de quantificar o quão ruim foi a tentativa. O processo de otimização é o método no qual o algoritmo checa os erros e então atualiza como o processo de escolha chega à decisão final para que evite errar da próxima vez.

Obtendo-se o breve conhecimento sobre os métodos de aprendizado e treinamento de IAs através de aprendizado de máquina, é possível obter-se a conclusão de sua importância. Devido à natureza de reconhecimento de padrões e categorização de dados, machine learning é recomendado para lidar com larga escala de dados e para encontrar possíveis diferenças de padrões que de outra forma poderiam escapar o olho e mente humana, ajudando em sistemas desde recomendações de mídia para consumir até descoberta de tumores em pacientes com predisposição ao câncer (UC Berkeley, 2020).

2.4. DEEP LEARNING

A IBM Cloud Education (2020) afirma que “o deep learning tenta imitar o cérebro humano — embora longe de corresponder à sua capacidade — permitindo que os sistemas agrupam dados e façam previsões com incrível precisão”.

Composta de uma rede neural com três ou mais camadas, deep learning é uma forma de machine learning não-supervisionado que recebe dados sem categorização ou agrupamento e então é responsável por criar seus próprios padrões de forma independente, sem intervenção de usuários (IBM CLOUD EDUCATION, 2020).

Um dos problemas observados com deep learning, como aponta IBM Cloud Education (2020), é o requisito de hardware necessário para execução do algoritmo de forma eficiente. Ao lidar com dados em larga escala, o tempo é a variável determinante para a escolha de tecnologia. Assim sendo, algoritmos que empenham deep learning se beneficiam de componentes com maior velocidade e quantidade de memória, sendo assim, unidades de processamento gráficas (GPU), comumente conhecidas como placas de vídeo, são preferidas como unidades de processamento devido sua capacidade de lidar com alta densidade de cálculos e velocidades de barramento.

2.5. REDES NEURAIS

Conforme IBM Cloud Education (2020), “As redes neurais refletem o comportamento do cérebro humano, permitindo que programas de computador reconheçam padrões e resolvam problemas comuns nos campos de IA, machine learning e deep learning". Segundo eles, redes neurais, ou neural networks, consistem em algoritmos que tentam imitar o funcionamento do cérebro humano, aprendendo a reconhecer padrões, se adaptando a novas situações e formando uma “rede de neurônios” que seguem aprendendo com cada novo processamento.

As redes neurais são utilizadas para o método de aprendizado deep learning, de forma que não é necessária interação manual humana. A rede neural é formada por uma ou mais camadas e os dados são processados por cada camada e então combinados para no fim obter um resultado preciso (AMAZON, 2022).

2.6. TECNOLOGIAS

Para melhor entender os principais agentes inteligentes presentes para uso público e privado, os mesmos serão divididos em duas categorias nesta seção:

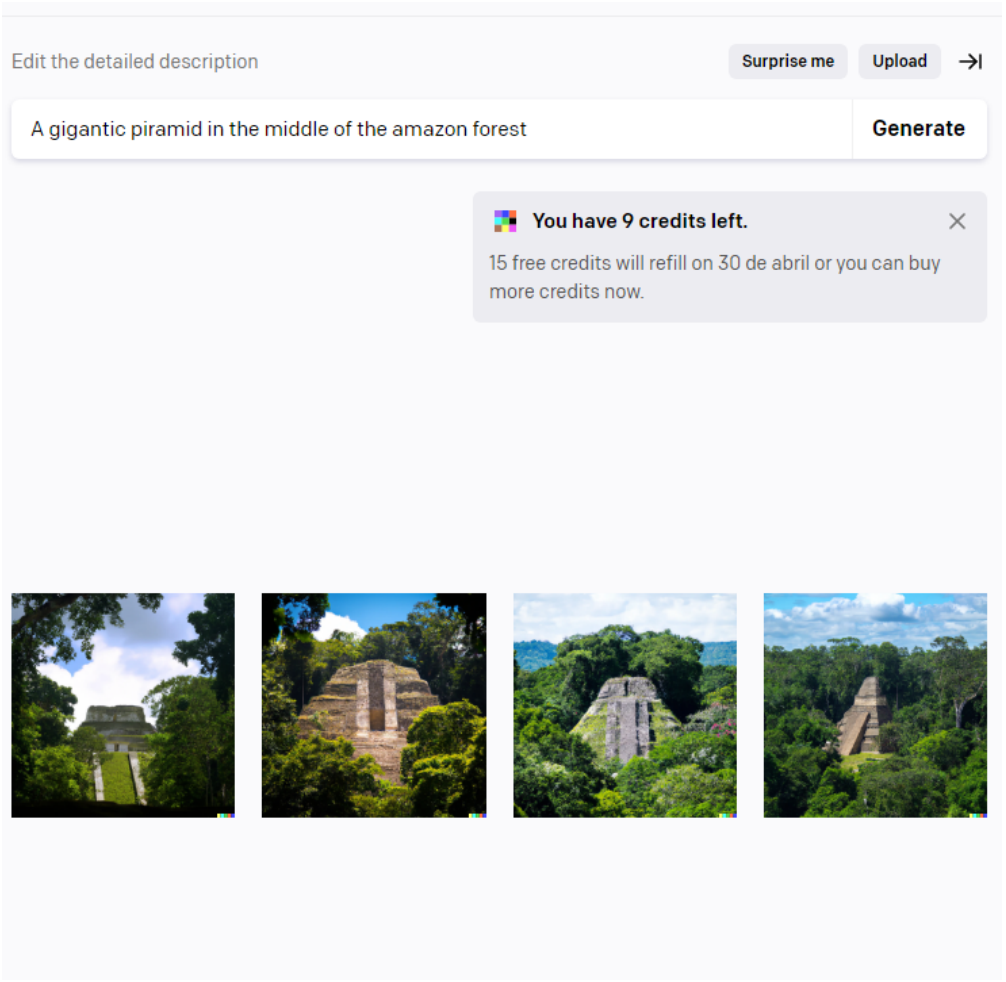
- Sistemas de manipulação e criação de informações;
- Sistemas de pesquisa e apoio.

2.6.1. **Dall-E**

Segundo o laboratório de pesquisa OpenAI (2021), “DALL·E é uma versão de 12 bilhões de parâmetros do GPT-3 treinado para gerar imagens a partir de descrições de texto, usando um conjunto de dados de pares texto-imagem”. Para utilizar o software, o usuário descreve em formato de texto como é a imagem que deseja, como mostra a figura 2. A

descrição pode ser feita com a utilização de frases semelhantes a conversações humanas e palavras-chave podem ser utilizadas opcionalmente para escolher formato e estilo artístico.

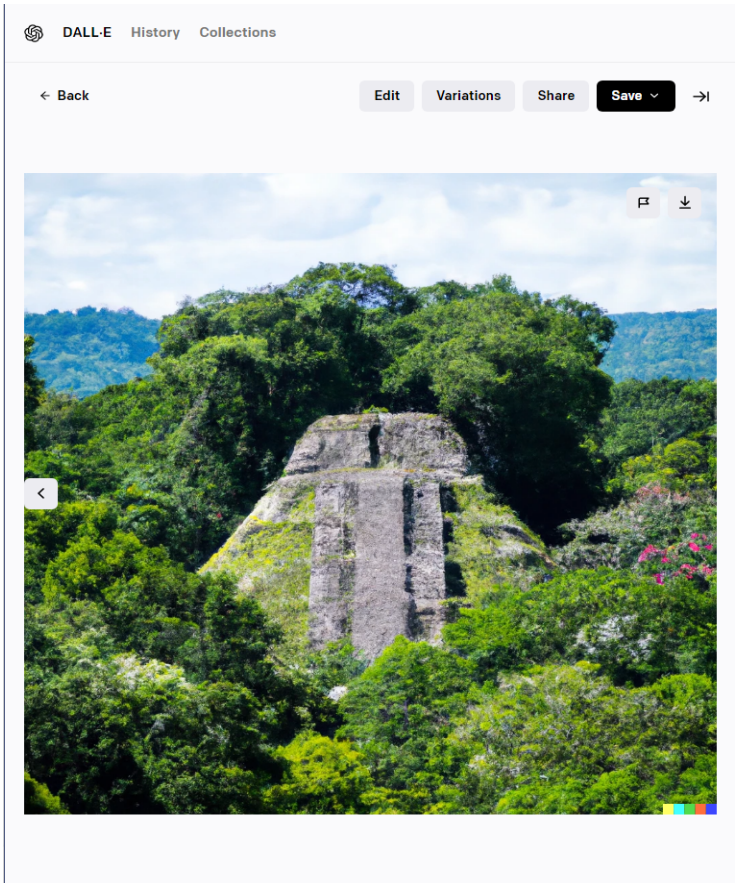
Figura 2 - Interface, prompt de texto e imagens geradas a partir de texto com DALL-E



Fonte: Autores (2023, online).

O software gera as imagens em grupos de 4 a 5 imagens por prompt. O usuário pode clicar em uma das imagens geradas para selecionar uma das imagens. Ao escolher a arte, opções para editar, obter variações, compartilhar e salvar são apresentadas como mostra a figura 3.

Figura 3 - Interface de usuário do Dall-E com botões para manipulação da imagem



Fonte: Autores (2023, online).

Semelhante ao modelo GPT-3, modelo transformador de texto também criado pelo laboratório OpenAI, DALL-E recebe tanto texto quanto imagens como um único fluxo de dados contendo até 1280 “tokens”. Os tokens, que são utilizados no treinamento do modelo, são utilizados com base na máxima probabilidade de gerar todos os tokens, um após o outro. Este método de treinamento não só permite o DALL-E a gerar imagens, mas também a regenerar qualquer região retangular de uma imagem que se estenda para a parte inferior direita de um jeito que seja consistente com o texto fornecido (OPENAI, 2021).

Alguns exemplos de criações usando a ferramenta incluem a figura 4, que mostra um cão e um pássaro tomando chá juntos no estilo de watercolor surrealista e a figura 5 que representa dois ursos de pelúcia fechando negócio em uma cidade cyberpunk iluminada por neon.

Figura 4 - Imagem criada usando o DALL-E com o prompt “cão e pássaro tomando chá ao estilo surrealista em watercolor”

Fonte: Autores (2023, online).

Figura 5 - Imagem criada usando o DALL-E com o prompt “dois ursos de pelúcia fechando negócio em uma cidade cyberpunk iluminada por neon.”

Fonte: Autores (2023, online).

A qualidade das imagens dependem em boa parte do prompt do usuário e a utilização de palavras chave, com termos como fotorrealista ou ao estilo de um artista determinando como a composição das imagens podem ser criadas. Por maior que sejam as discrepâncias entre as duas figuras, ambas foram criadas pela mesma aplicação, na mesma interface.

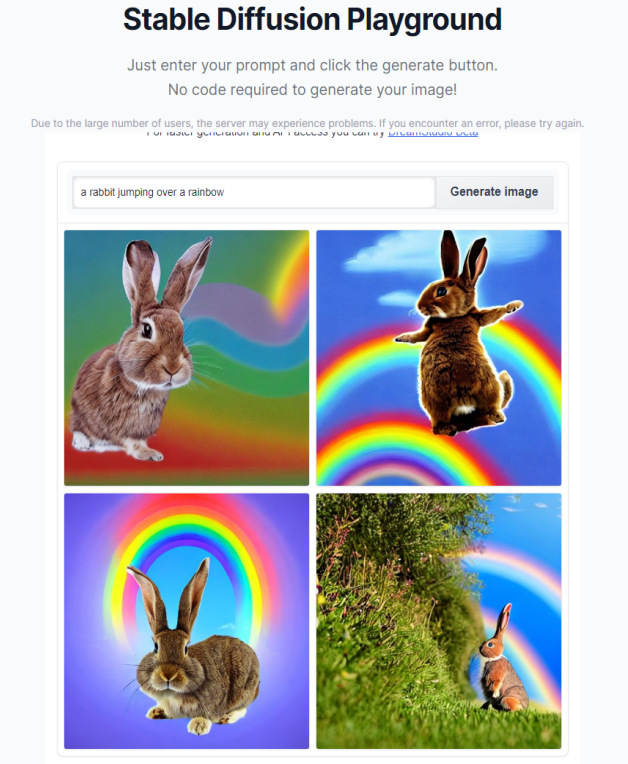
2.6.2. Stable Diffusion

Stable Diffusion é uma Inteligência Artificial desenvolvida pelo grupo Stability AI e disponibilizada em 2022, utilizando deep learning de difusão para gerar imagens a partir de prompts de texto, apresentando semelhanças com seu concorrente Midjourney (STABILITY AI, 2022).

A IA foi treinada com base em dados coletados por um *web crawler* chamado LAION-5B, que rastreia informações disponíveis na internet, sendo popular no treinamento de modelos generativos de diversas empresas no ramo (LAION, 2022). No entanto, ao comparar com IAs concorrentes, o Stable Diffusion vai além pois permite a remoção e inclusão de elementos em imagens já existentes.

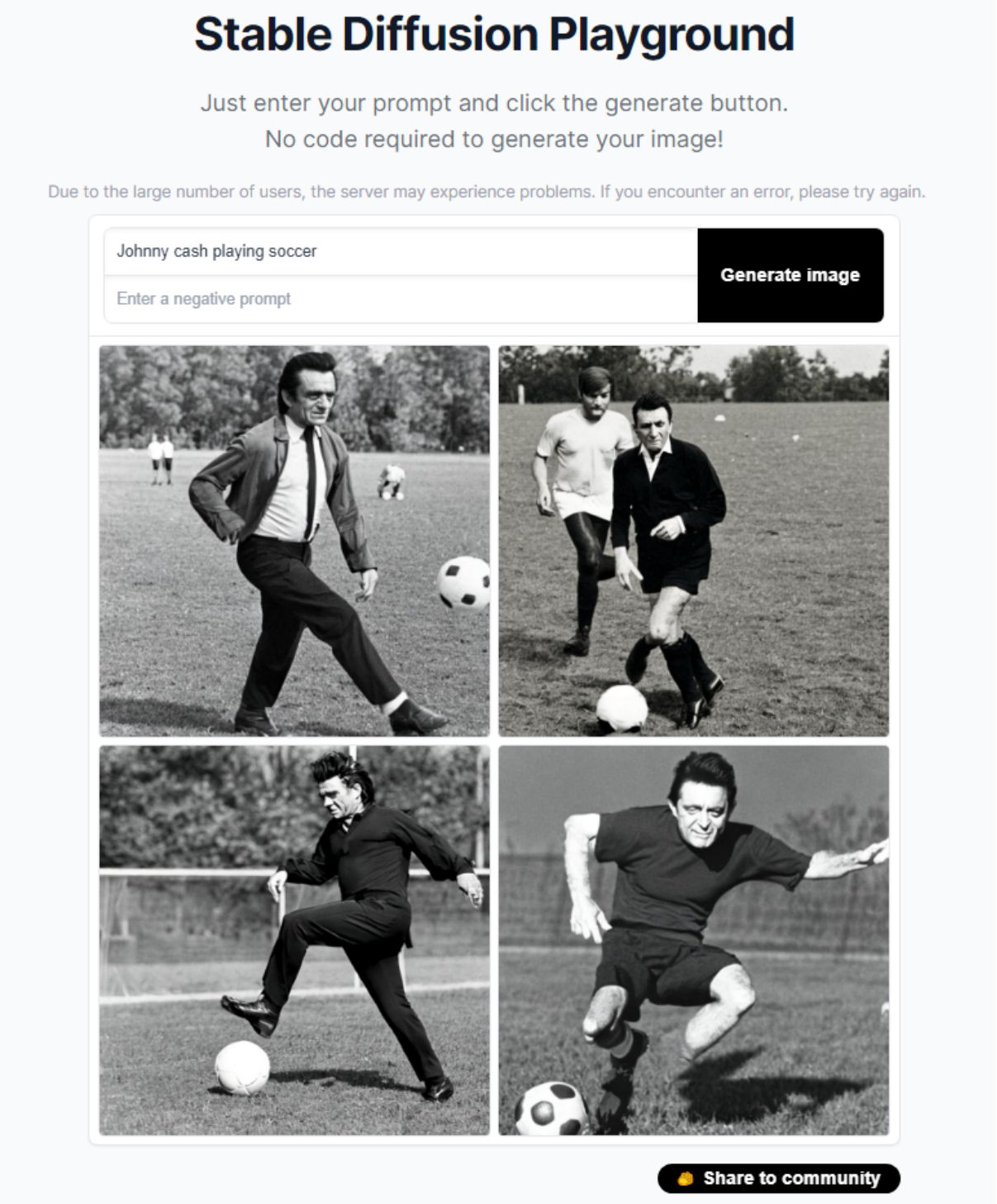
As criações do Stable Diffusion impressionam devido sua qualidade, como mostra a Figura 4, representando um coelho saltando sobre um arco-íris, gerado na demonstração online do Stable Diffusion.

Figura 6 - Imagens geradas via demo online do Stable Diffusion, com o prompt "coelho pulando sobre um arco-íris".



Fonte: Autores (2023, online).

Figura 7 - Imagens geradas via demo online do Stable Diffusion, com o prompt "Johnny Cash jogado Futebol".



Fonte: Autores (2023, online).

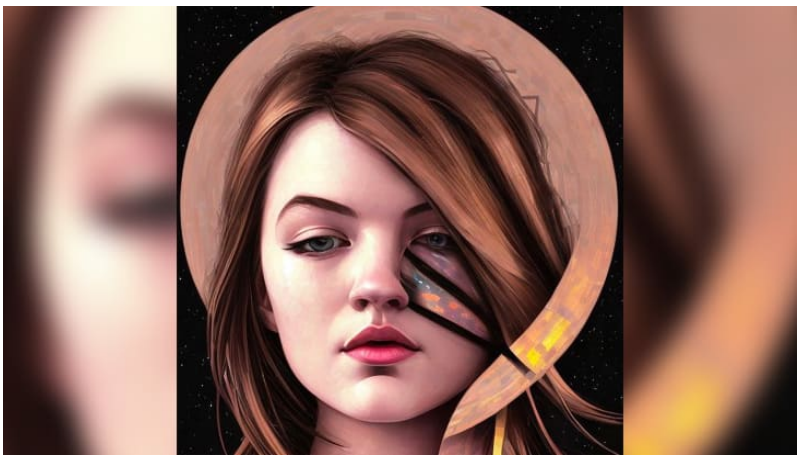
Diferente do Midjourney e do DALL-E, o Stable Diffusion é um software de código aberto e devido a isto, estão disponíveis a opção de baixar o código-fonte e os modelos, possibilitando a criação de imagens diretamente no computador do usuário.

2.6.3. Lensa AI

Criado pela empresa Prisma, Lensa AI é um aplicativo móvel pago com sistema de compras e assinatura dentro do próprio software. De acordo com os termos de privacidade do aplicativo, Lensa AI utiliza o modelo Stable Diffusion e para treiná-lo, utiliza o banco de dados LAION-5B, formado por uma enorme rede de arte digital extraída da internet (CNN, 2022).

O aplicativo, disponível para dispositivos com sistemas operacionais Android e iOS, apresenta funções como editor de fotos e os *"magic avatars"*, retratos gerados através do *"Stable Diffusion"* em estilos artísticos populares, como o retrato em estilo arte digital de Zoe Sottlie, mostrado na Figura 5. Os avatares são gerados a partir de fotos de rosto *"selfie"* fornecidas pelo usuário. A função de *"magic avatar"* é vendido separadamente da assinatura do aplicativo e possui valor por número de geração de avatares, com opção de 50, 100 e 200 avatares gerados.

Figura 8 - *"Magic Avatar"* da redatora Zoe Sottlie, CNN, utilizando o Lensa.



Fonte: Zoe Sottlie, Lensa (2023, online).

Segundo a Prisma Labs (2023) em seus termos de privacidade, o aplicativo pode utilizar da tecnologia TrueDepth API, desenvolvida pela Apple, para determinar dados como posição e topologia do rosto no processo de geração de *"magic avatars"* ocorre nos seguintes passos:

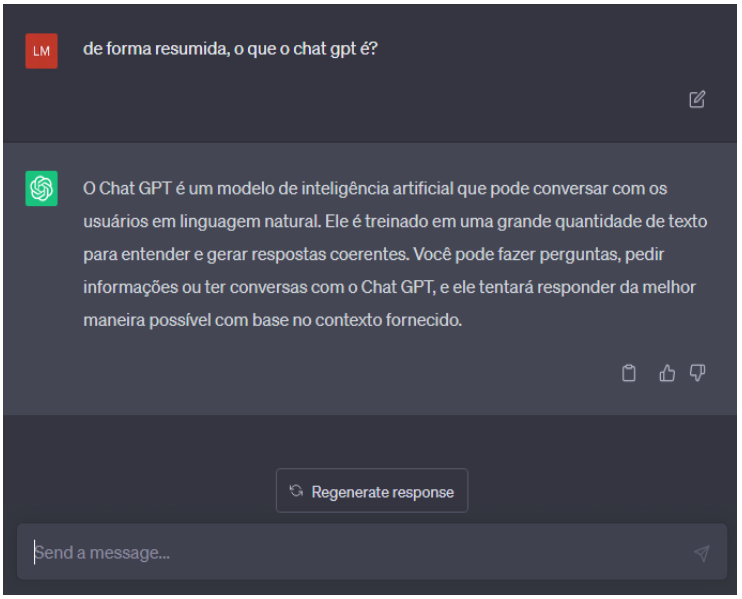
1. O usuário faz upload de 10 a 20 imagens e seleciona o gênero que se identifica;

2. Usando os dados fornecidos, uma cópia do modelo *"Stable Diffusion"* é re-treinado para personalizar o modelo;
3. Após a cópia do modelo ser re-treinado os retratos são gerados;
4. Após os retratos serem gerados, a cópia do modelo junto com os dados fornecidos é deletada permanentemente dos servidores. Os *"magic avatars"* comprados são salvos permanentemente nos servidores para fácil acesso no futuro.

2.6.4. Chat-GPT

O ChatGPT se trata de um modelo de I.A criado pela OpenAI que utiliza o modelo de linguagem autoregressivo conhecido GPT-3.5, que se trata da combinação do modelo GPT 3, que é a terceira geração do modelo de linguagem generativa e autoregressiva GPT pré treinado utilizando Deep Learning utilizando a metodologia RLHF, que em uma tradução livre significa "Aprendizado reforçado por feedback humano", onde o modelo tem a possibilidade de se aperfeiçoar de acordo com o feedback do usuário, baseado num sistema de recompensa (OPENAI, 2022).

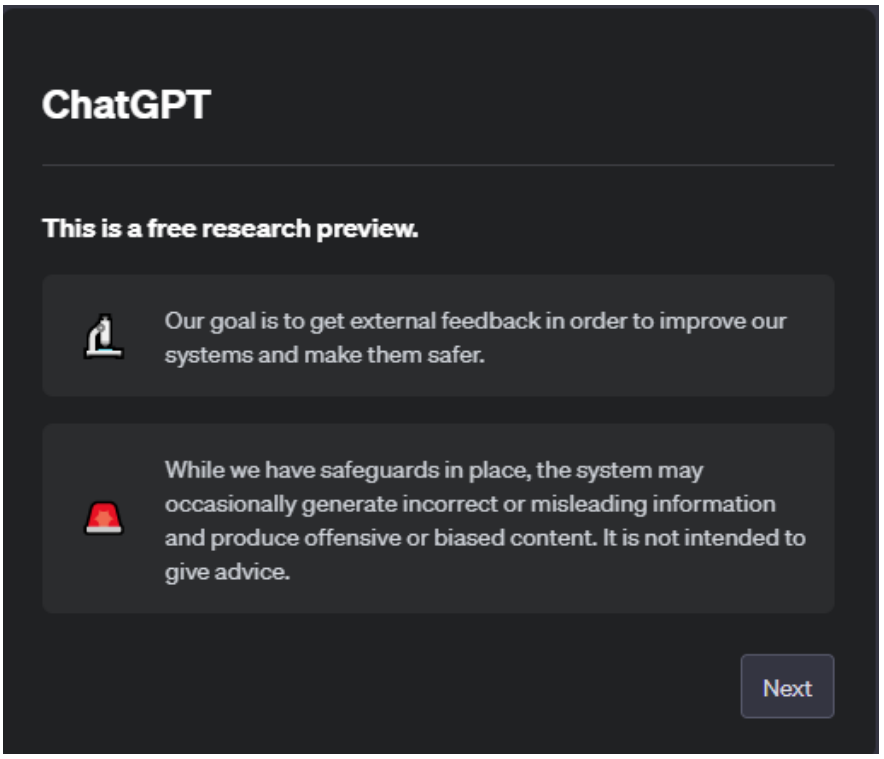
Figura 9 - ChatGPT respondendo uma pergunta sobre "O que é ChatGPT?"



Fonte: Autores (2023, online).

Durante a produção deste trabalho, a ferramenta está disponível para qualquer pessoa que se cadastrar no site da OpenAI, onde é possível testar a aplicação em tempo real.

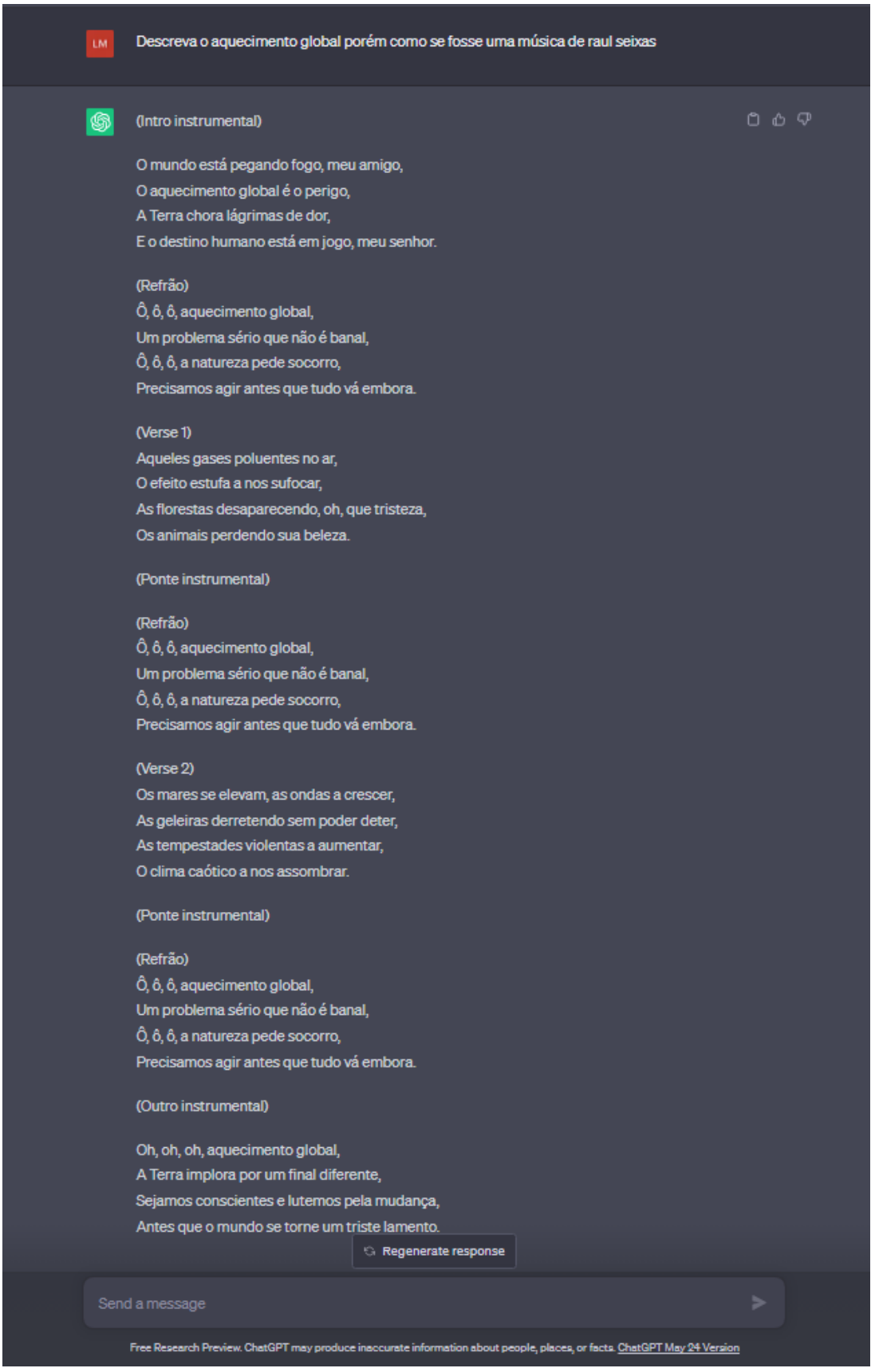
Figura 10 - ChatGPT avisando das possibilidades de informações incorretas, com viés ou enganosas



Fonte: Autores (2023, online).

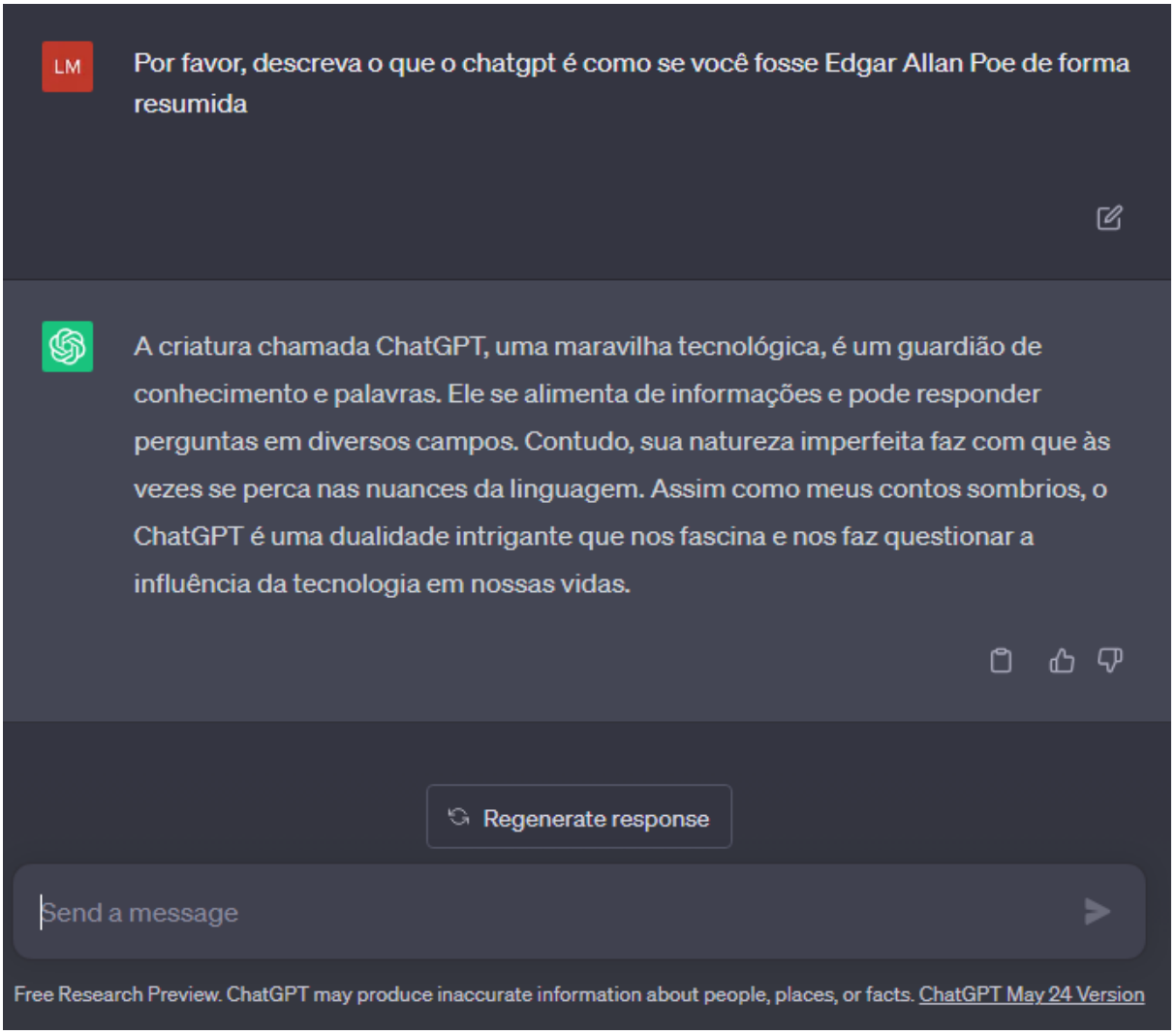
Há outros exemplos interessantes de uso da ferramenta, que podem ser encontrados nas figuras 10 e 11. Na figura 10, foi solicitado ao ChatGPT que descrevesse o aquecimento global no estilo característico das músicas do icônico cantor brasileiro Raul Seixas. Já na figura 11, foi pedido ao ChatGPT que explicasse o que é o ChatGPT utilizando o estilo de escrita do renomado poeta e autor Edgar Allan Poe.

Figura 11 - ChatGPT descrevendo o aquecimento global como uma música de Raul Seixas



Fonte: Autores (2023, online).

Figura 12- ChatGPT descrevendo a si mesmo no estilo de Edgar Allan Poe



Fonte: Autores (2023, online).

Vale ressaltar que por ser uma ferramenta ainda em desenvolvimento, a mesma no momento de escrita desse trabalho era capaz de fornecer informações erradas ou incompletas, mesmo que os dados já estejam presentes em seu modelo.

2.6.5. **Midjourney**

Midjourney trata-se de uma inteligência artificial focada na geração de imagens baseadas em prompts de texto fornecidas pelo usuário (ARIMETRICS, 2022), de forma semelhante aos concorrentes DALL-E e Stable Diffusion.

Figura 13 - Imagem gerada pelo Midjourney para o prompt “uma casa feita de cogumelos amanita muscaria realista”

Fonte: Autores (2023, online).

Durante a escrita deste trabalho, a única forma de utilizar o programa é via uma comunidade no aplicativo Discord, onde o Midjourney requer uma inscrição paga para gerar imagens, que são feitas no canal de mensagens do discord por meio de comandos como por exemplo “/imagine” (MIDJOURNEY, 2022), que gera imagens de acordo com o prompt inserido pelo usuário.

Figura 14 - IA Midjourney disponibilizada via Discord

Fonte: Autores (2023, online).

Tendo entrado em versão beta aberta para usuários em 12 de julho de 2022 (MIDJOURNEY, 2022), o Midjourney tornou-se uma das I.As mais populares no ano de seu lançamento devido sua facilidade de uso, fácil acesso em múltiplas plataformas, como o Discord, mostrado na Figura 9, e interpretação de *prompts* de texto criando resultados de alta qualidade, principalmente comparados aos resultados da concorrente DALL-E.

Por conta da qualidade elevada e realismo que pode ser alcançado pelas imagens geradas, o modelo generativo possui limitações que impedem *prompts* com conteúdos sexuais ou envolvendo violência "gore" (Termo usado para mídias com Violência Extrema).

2.6.6. **Deepfake**

O termo Deepfake não se refere a um software específico mas sim a mídia sintética criada com a utilização de agentes inteligentes que empregam o método de treinamento deep learning para criar vídeo, áudio e outras formas de mídia utilizando características pessoais de alguém (BUSINESS INSIDER, 2023).

Segundo Somers (2020), o termo Deepfake se tornou popular quando, no final de 2017, um usuário com o nome Deepfake postou vídeos pornográficos que utilizavam tecnologia de código aberto para troca de rostos na rede social Reddit. Desde então, tecnologias antigas e novas que criam mídia sintética acabaram por ser categorizadas como Deepfakes.

Devido à conotação negativa ligada ao termo, cada vez mais pessoas no ambiente de negócios referem-se a mídia sintética utilizada para finalidades como publicidade e marketing como "mídia sintética gerada por inteligência artificial", termo amplo o suficiente para incluir a definição original de Deepfakes mas específico o suficiente para omitir outras formas de alteração de mídia que não envolvem IA, como imagens de computação gráfica em filmes ou imagens editadas no photoshop, por exemplo (SOMERS, 2020).

Para criar um vídeo Deepfake, é necessário a utilização de um algoritmo de reconhecimento facial e uma rede computacional de Deep Learning chamado VAE (*Variational Auto-Enconder*). VAEs são treinados para codificar imagens em representações de baixa dimensão e depois decodificá-las novamente em imagens (SOMERS, 2020).

Somers (2020) explica que para trocar a face de um ator por outro em um vídeo, por exemplo, é necessário a utilização de dois VAEs. O primeiro é treinado com imagens da face do ator que substituirá o rosto, o segundo é treinado com imagens de uma grande variedade de faces. Quando o treinamento é concluído, o codificador das várias faces é combinado com o decodificador das imagens da face do ator, resultando no rosto do mesmo no corpo de outra pessoa.

2.7. APLICAÇÕES DA INTELIGÊNCIA ARTIFICIAL NA MEDICINA

No artigo Artificial Intelligence In Medicine, AN Ramesh, C Kambhampati et. al. (2004) analisam sobre a utilização e pesquisa da inteligência artificial no campo da medicina e especificamente na área de diagnóstico e prognóstico de doenças como câncer.

Um dos principais pontos a se notar é a menção das redes neurais artificiais e sua grande vantagem na área da medicina. De acordo com AN Ramesh, C Kambhampati et. al. (2004), “a habilidade que redes neurais possuem de aprender a partir de exemplos históricos, analisar dados não lineares, lidar com informação imprecisa e generalizar, permitindo a aplicação do modelo a dados independentes tornou as redes neurais a ferramenta analítica mais atrativa no campo da medicina”.

Na área de diagnósticos, AN Ramesh, C Kambhampati et. al. (2004) falam sobre o algoritmo ProstAsure Index. Este, um algoritmo de classificação derivado de rede neural, consegue classificar próstatas entre benignas e malignas. No artigo, os autores mencionam que o modelo foi subsequentemente validado por estudos e possui precisão de diagnóstico de 90%, com uma sensibilidade de 81% e especificidade de 92%.

Além da utilização no diagnóstico de câncer de próstata, redes neurais artificiais também foram utilizadas em outros diagnósticos cirúrgicos relevantes como dores abdominais e apendicite, cálculo biliar, glaucoma e dores nas costas (AN RAMES, C KAMBHAMPATI ET AL., 2004).

Além de diagnósticos, redes neurais artificiais também foram utilizadas no prognóstico de pacientes. Em casos de doenças onde tratamentos agressivos podem colocar a vida do paciente em risco, redes neurais artificiais apresentaram-se de extrema utilidade. Segundo AN Ramesh, C Kambhampati et. al. (2004), modelos podem ser treinados para

calcular a probabilidade de sobrevivência de um paciente a tratamentos agressivos e de alto risco como quimioterapia e radioterapia para pacientes com câncer. Desta forma, decisões sobre qual tratamento utilizar para casos específicos de câncer pode aumentar as chances de sobrevivência.

Apesar do grande potencial da inteligência artificial em contribuir com diagnósticos e prognósticos, Buch, Ahmed e Maruthappu (2018) afirmam que a tecnologia não irá substituir médicos humanos, mas sim auxiliá-los com tarefas repetitivas como extrair dados de pacientes e identificar riscos com base em históricos.

2.8. APLICAÇÕES DA INTELIGÊNCIA ARTIFICIAL NA EDUCAÇÃO

A utilização da inteligência artificial na educação é referida como AIED (*Artificial Intelligence in Education*, ou Inteligência Artificial na Educação em português). Segundo Bialik, Fadel & Holmes (2019), a AIED divide-se em duas categorias: IA para apoio no aprendizado e IA para entendimento do aprendizado. IA para apoio no aprendizado refere-se a ferramentas que auxiliam alunos e professores na retenção e transmissão de conhecimento. IA para entendimento do aprendizado, refere-se a ferramentas que analisam mecanismos e métodos de estudo e transmissão de conhecimento além de como ocorreram.

Com o crescimento da IA em diversas áreas da economia, os investimentos na AIED aumentam de forma exponencial. Bialik, Fadel e Holmes (2019) afirmam que até o ano de 2024 o valor de mercado da ciência da IA na educação será de aproximadamente US$ 6 bilhões.

De forma mais específica, um dos grandes avanços da IAED encontra-se no sistema chamado ITS (*Intelligent Tutoring System*, ou sistema de tutoragem inteligente em português). Segundo Chen et al. (2022), os ITS se tratam de sistemas cuja função é proporcionar aos estudantes uma experiência de aprendizagem mais adaptada para cada indivíduo, por meio dos conhecimentos que o ITS obteve de um professor ou expert somado aos erros e acertos dos alunos. Utilizando-se do modelo de aprendizado de IA supervisionado, os ITS demonstram grandes benefícios para estudantes com dificuldades de aprendizado. Chen et al. (2022) ressalta os resultados positivos no aprendizado para estudantes no espectro autista devido a rapidez e especificidade do feedback que o sistema retorna ao aluno.

Além dos ITS, que desenvolvem planos e sistemas de ensino específicos para cada indivíduo, Chen et al. (2022) citam a importância do ISLA (*Integrative Specialized Learning Application* ou aplicação integrativa de aprendizagem especializada), um sistema que utiliza IA e possibilita que as emoções dos alunos sejam levadas em consideração ao dar feedback e encorajamento. Em suas pesquisas, Chen et al. (2022) mencionam sobre a importância do estado emocional do estudante em relação a performance de aprendizado, principalmente com estudantes do espectro autista.

Mousavinasab et al. (2018) listam as principais utilizações de sistemas ITS e frequência de uso baseado na coleta de dados sobre AIED. Notavelmente, a IA aplicada em ITS é mais comumente utilizada para fornecer feedback adaptativo, definir, ou atualizar as características do estudante e avaliação do estudante, como mostra a tabela:

Tabela 2 - Frequência da utilização de IA em sistemas ITS

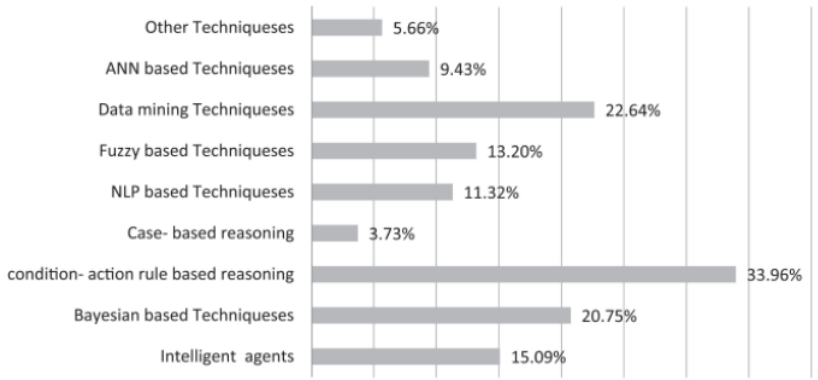
Table 2. The purposes of applying AI techniques in ITSs.

The purposes of applying AI techniques in ITSs	Frequency (%)
Adaptive feedback, hint or recommendation generation	52.83%
Defining, classification, or updating the learner's characteristics	56.60%
Learner's evaluation	45.28%
Presenting adaptive learning material or content	41.50%
Adaptive learning path navigation	28.30%
Presenting adaptive test and exercises	5.66%

Fonte: Mousavisanab et al. (2018, p.15)

Diversas técnicas de aprendizado e implementação de inteligência artificial são utilizadas na educação, com sua maioria utilizando sistemas de raciocínio baseado em regra de condição - ação, técnicas de mineração de dados e técnicas baseadas em inferência Bayesiana como mostra o gráfico (MOUSAVISANAB ET AL., 2018):

Gráfico 1 - Parcelas das áreas de estudo onde IA é implementada na educação



Fonte: Mousavisanab et al. (2018, p. 15).

Como apontado no Gráfico 1, as três maiores áreas de estudo com onde há implementação de IA na educação são, em ordem decrescente, “Condition-action rule based reasoning” que em uma tradução livre significa “Raciocínio baseado em regras de condição e ação” definido por Berndtsson e Mellin (2009) como a sequência de ocorrência de um evento, análise do mesmo com uma condição e tomada de decisão com base nessa condição, “Data mining techniques” que se traduz como “Técnicas de mineração de dados” que engloba métodos de mineração de dados comumente utilizados por grandes empresas para determinar resultados com base em dados analisados, e por fim “Bayesian based techniques” que em portugûes significa “Técnicas baseadas em Bayesian”, que são métodos estatísticos baseados no teorema matemático criado por Thomas Bayes, que como apontado por Britannica (2023) pode ser definido como combinação de uma população de eventos conhecida com evidências tiradas de uma amostra para auxiliar na inferência estatística.

2.9. APLICAÇÕES DA INTELIGÊNCIA ARTIFICIAL NA SEGURANÇA

A utilização de IA no espaço civil vai além da educação e saúde, mas também na área de segurança. Srivastava, Bisht e Narayan (2017) citam em seu estudo o uso de detecção de sons de tiro na cidade de São Francisco utilizando o sistema chamado Shotspotter.

Em uma entrevista, o chefe de polícia de São Francisco, Gregory P. Suhr afirmou que o crescimento da violência armada acabou por dessensibilizar a população devido a frequência de ocorrências e com isso os números de denúncias de sons de tiros reduziram, por civis não quererem se envolver ou acharem que outras pessoas irão denunciar. Com a utilização do Shotspotter, o tempo de resposta da polícia contra casos de violência armada se tornou mais rápida e consequentemente a taxa de homicídio devido a uso de armas de fogo caíram (SRIVASTAVA; BISHT; NARAYAN, 2017).

Utilizando reconhecimento de padrão temporal, o Shotspotter é capaz de discernir entre sons de tiros e outros sons de explosões contidas como carros tendo dificuldade de iniciar seu motor. Além da capacidade de identificar sons de tiros, o sistema que utiliza uma rede neural artificial (ANN) é treinado utilizando uma técnica de filtro espacial, permitindo que aprenda a identificar o perfil sonoro de tiros até mesmo em ambientes poluídos

sonoramente. Ainda segundo Srivastava, Bisht e Narayan (2017), o sistema não é limitado apenas a ajudar com resposta policial, mas também auxilia a promotores e cientistas forenses com sua capacidade de identificar qualidade do tiro, espaço de tempo entre eles, local e hora exatas do acontecimento.

3. MÉTODO

Este capítulo tem como objetivo descrever os métodos de pesquisa utilizados para a criação deste artigo, as etapas da pesquisa, bem como seu tipo e também o escopo no qual se insere.

3.1. TIPO DE PESQUISA

A pesquisa é definida por Creswell, J.W (2008) como "[...] um processo de passos utilizados para coletar e analisar informação para aumentar o nosso conhecimento de um tópico ou problema". Além desta definição, Creswell J.W (2008) divide a pesquisa em 3 partes, listadas como:

1. Propor uma questão a se analisar e responder.
2. Coletar dados de forma que seja possível responder a questão.
3. Apresentar a resposta para a questão proposta inicialmente.

Seguindo-se estes passos, o trabalho possui como objetivo expor as múltiplas faces da utilização da IA, suas contribuições positivas e negativas, e seu impacto na sociedade, a partir de dados coletados por meio de pesquisas em artigos, sites, livros e documentos.

A pesquisa bibliográfica é a revisão da literatura sobre as principais teorias que norteiam o trabalho científico. Essa revisão pode ser feita em diversas fontes em diferentes mídias como livros, periódicos, artigos de jornais, sites da Internet e outras fontes. (PIZZANI; SILVA; BELLO; HAYASHI, 2012). Através da pesquisa e revisão de obras de outros autores é possível obter padrões e noções de potencial da tecnologia, ajudando a fazer previsões embasadas sobre o possível futuro da inteligência artificial.

Sendo assim, a pesquisa bibliográfica é o método de pesquisa escolhido para o trabalho devido ao amplo volume de trabalhos científicos que abordam o tema disponíveis para estudo, possibilitando a oportunidade de agregar, estudar e compilar tais trabalhos em uma visão mais ampla e rica do assunto.

3.2. ETAPAS UTILIZADAS

As etapas do processo de elaboração deste projeto são quantificadas em:

1. **Estudo sobre a trajetória do campo da inteligência artificial** desde sua criação, consistindo de pesquisas de artigos, notícias e dados sobre a origem e evolução da inteligência artificial até o momento atual da história a fim de contextualizar o assunto e demonstrar o progresso até então obtido pela área.
2. **Estudo sobre o cenário contemporâneo da inteligência artificial** no mundo globalizado, sendo necessário estudar dados de destaque relacionados à tecnologia no período entre os anos de 2022 e 2023 tendo como exemplo, mas não limitado a áreas de aplicação mais utilizadas, informações disponibilizadas em notícias e jornais e publicações científicas sobre o assunto.
3. **Compilar, formatar e separar os dados reunidos** organizando as informações em ordem de relevância tanto midiática como científica, extraindo os dados úteis para o uso no trabalho e por fim separá-los por tópicos para melhor estruturação.
4. **Identificar os pontos de maior destaque e repercussão** combinando os tópicos anteriormente separados a fim de discutir possíveis perigos e criticidades do tema.
5. **Levantar uma discussão** sobre a criação e utilização da inteligência artificial, apresentando opiniões e expondo pontos importantes sobre o uso ético da tecnologia

3.3. DELIMITAÇÕES

O projeto se foca na área da inteligência artificial, com o intuito de obter uma discussão direta e concisa evitando desvios do assunto, salvo casos onde o mesmo seja necessário para auxiliar na contextualização de um ponto abordado.

O projeto em questão utiliza-se de dados nacionais e internacionais, em português e inglês com a finalidade de formar uma base de dados maior, criando uma visão mais abrangente sobre os efeitos da tecnologia no escopo do país e do planeta.

4. DISCUSSÕES SOBRE O USO DA INTELIGÊNCIA ARTIFICIAL GENERATIVA

A inteligência artificial apresenta impactos muito maiores para a sociedade do que é comumente imaginado. Da utilização para melhorar diagnósticos médicos a criação de mídia sintética para fins políticos, a IA é uma ferramenta moralmente neutra cuja finalidade é determinada pelo utilizador.

Este capítulo tem como intuito apresentar e discutir o impacto da IA na sociedade e discutir de forma crítica sua utilização e possíveis formas de coibir os problemas que podem surgir com sua utilização.

De acordo com relatório da empresa Deeptrace, organização especializada em detecção de Deepfakes, em setembro de 2019, 96% dos Deepfakes encontrados online são de conteúdo pornográfico. Em sua maioria, estes conteúdos são criados sem consentimento da vítima e apresentam a mesma de forma explícita causando danos à reputação e possíveis outros problemas (FORBES, 2020).

Segundo Carvalho (2022), os Deepfakes podem ser punidos de acordo com a legislação brasileira, apesar das leis que englobam o assunto não serem feitas especificamente para crimes envolvendo mídia sintética. Para fins sexuais, a lei 11.829/08, de 2008, pune a montagem envolvendo menores de idade em cenas de sexo explícito ou pornográfico (artigo 241-C do ECA). Já em 2018, o artigo 216-B do Código Penal passou a punir montagens pornográficas com maiores de idade, crime punível de 6 meses a 1 ano e multa.

Os Deepfakes parecem problemas isolados, que afetam apenas famosos, mas a facilidade da utilização de redes neurais de deep learning em computadores caseiros ou serviços de computação em nuvem e o fácil acesso a fotos e vídeos em redes sociais para treinamento destas redes neurais, cria risco real de qualquer pessoa se tornar uma vítima de mídias sintéticas com fim de injúria, difamação, calúnia ou até como criação de conteúdo pornográfico. Aliado a facilidade de criação, a velocidade com que a mídia é compartilhada online e a dificuldade de desmentir “provas sólidas”, como vídeos, gravações de áudio e afins, cria uma difícil batalha contra avanços tecnológicos e sua amoralidade.

A utilização imoral de Deepfakes não se limita somente a ataques pessoais. Segundo notícia reportada por Catherine Stupp no website The Wall Street Journal, em 2019, um caso de estelionato ocorreu contra uma firma de energia baseada no Reino Unido. O CEO da filial achou que estava falando com o chefe executivo da matriz da empresa, localizada na

Alemanha. Na ligação, o suposto chefe executivo demanda que £ 220.000,00 fossem transferidos para a conta de um fornecedor húngaro. O ligador exigiu que a transferência acontecesse dentro de 1 hora, de acordo com a empresa de seguros da firma, Grupo Euler Hermes S/A. Os hackers ligaram novamente uma segunda vez para dizer que a transferência havia sido reembolsada pela matriz e então uma terceira vez, de um número austríaco, demandando outro pagamento (STUPP, 2019).

Em Stupp (2019), é comentado que "'O CEO britânico reconheceu o leve sotaque alemão de seu chefe e a melodia de sua voz pelo telefone', disse Rüdiger Kirsch, um especialista em fraudes na Euler Hermes".

O crime ocorrido em 2019 demonstra o perigo de mídias sintéticas quando utilizadas para enganar pessoas específicas, principalmente ao empregar sintetização de voz para enganar pessoas, tornando a engenharia social cada vez mais difícil de detectar e evitar. No caso da firma britânica não nomeada, a demora para o reembolso junto ao telefone de um país desconhecido foi o que fez o CEO suspeitar do caso e não fazer uma segunda transferência.

O que torna mais assustador a utilização de Deepfakes de voz para estelionato é que as principais vítimas de golpes no Brasil podem se tornar alvos desta técnica. Com áudios e vídeos sendo enviados diariamente em grupos de mensagem instantânea e redes sociais, dados para treinamento de redes neurais deep learning se tornam cada vez mais fáceis de obter. Um criminoso com informações sobre uma vítima que possui dinheiro em sua conta de poupança, por exemplo, pode treinar uma IA para modificar sua voz e soar como um parente de confiança da vítima, enganando-a a transferir dinheiro para sua conta.

Uma ameaça ainda maior que estelionatos e difamação é a utilização de Deepfakes no âmbito político. Em 2018, um vídeo do ex-presidente dos Estados Unidos Donald Trump fazendo um discurso pedindo para que a Bélgica deixasse o Acordo de Paris foi divulgado por um partido político belga para progredir sua agenda. Trump, no entanto, nunca deu o discurso, sendo o vídeo uma mídia sintética criada com fins políticos (BURCHARD, 2018).

Para combater os problemas e perigos da desinformação causada por mídias sintéticas, o website Witness Media Lab foi criado pela Organização Não-Governamental Witness em 2015, com o intuito de discutir métodos para desmascarar Deepfakes e outras mídias sintéticas e apresentar dados e pesquisas relacionadas ao assunto. A ONG, que almeja utilizar mídia digital como "testemunhas" para documentação de fatos reais e ferramenta de

proteção aos direitos humanos, enfrenta desafios relacionados à autenticidade de mídias digitais (WITNESS, 2018).

Visando facilitar o acesso à informação, o projeto "*Prepare, Don't Panic: Synthetic Media and Deepfakes"* foi criado em 2018 e está disponibilizado no website Witness Media Lab, este apresenta maior foco na disponibilização de recursos, identificação e desmascaramento de deepfakes além de disponibilizar tutoriais de como criar mídia autêntica e verificável (WITNESS, 2018).

"In Event of Moon Disaster" é um projeto criado em 2019 por Francesca Panetta e Halsey Burgund, diretora de realidade estendida no MIT e artista membro do MIT Open Documentary Lab, respectivamente. Com o intuito de tornar mais conhecido o risco da desinformação causada pelo uso de Deepfakes, o projeto criou um vídeo de mídia sintética utilizando o discurso de resignação do ex-presidente dos Estados Unidos, Richard Nixon e o rascunho do discurso que seria dado caso os astronautas da Apollo 11 não tivessem sucesso ao retornar da Lua, feito pelo escritor de discursos de Nixon, Bill Safire (MIT OPEN LEARNING, 2019).

No vídeo, que mistura técnicas de Deepfake com substituição de lábios, adição de movimentos das mãos e olhos e outros detalhes de efeitos visuais, foi feito no início da pandemia do COVID-19 e possuía o objetivo de trazer luz ao perigo da desinformação e notícias falsas, afirma Somers (2020). O vídeo foi disponibilizado online com avisos de que a mídia foi sintetizada e gravações documentando o processo mostram como o projeto foi feito.

Semelhante a outras tecnologias, mídias sintéticas não são uma ferramenta exclusivamente maléfica. Thompson (2020) em seu artigo sobre o documentário independente *"Welcome to Chechnya"* ressalta o uso revolucionário de técnicas de Deepfake aliados a efeitos visuais para mostrar rostos humanos e expressões de emoção enquanto protege as identidades dos refugiados LGBTQIA+ que estavam a ser caçados, presos, torturados e até mesmo mortos na região da Chechênia.

Com intuito de proteger e melhorar a experiência online de jogadores, a Modulate, uma empresa com sede em Cambridge, Massachusetts, cria "cosméticos" de voz para utilização em jogos online e plataformas sociais. A Modulate disponibiliza uma coleção de vozes sintéticas que seus clientes, empresas no ramo de jogos online, podem integrar a suas plataformas, permitindo que os usuários modifiquem suas vozes (SOMERS, 2020).

O objetivo da tecnologia vai desde permitir uma maior imersão para o jogador, permitindo com que sua voz e a de seu avatar no jogo combinem da forma que imagina, até mesmo proteger alvos de assédio online como mulheres e crianças. Ao utilizar o modificador

de voz, o usuário pode se tornar mais anônimo e apenas revelar sua verdadeira identidade para pessoas que realmente confia.

Em janeiro de 2023, uma ação coletiva foi criada contra os criadores das ferramentas Midjourney, Stable Diffusion e a IA DreamUp, da DeviantArt. A ação foi criada por três artistas na área de arte digital, Sarah Andersen, Kelly McKernan e Karla Ortiz, e alega que os direitos dos artistas foram infringidos por conta da forma que estas ferramentas de arte generativas de IA são treinadas. Na ação coletiva, os artistas acusam as organizações administradoras das ferramentas de violarem os direitos autorais de milhões de artistas ao usar bilhões de artes coletadas da internet para treinar seus modelos neurais de geração de imagem (VINCENT, 2023).

A ação judicial está sendo movida pelo advogado Matthew Butterick junto ao escritório de advocacia Joseph Saveri. Butterick e Saveri também são responsáveis por uma ação judicial contra a Microsoft, Github e OpenAI devido ao programa GitHub Copilot, que utiliza código de programas open-source sem consentimento dos criadores para treinar a IA. Butterick (2023) descreve os casos como “mais um passo para tornar a IA justa e ética para todos” e ressalta que a capacidade da Stable Diffusion de “[...] inundar o mercado com um número essencialmente ilimitado de imagens infratoras causará danos permanentes no mercado de arte e artistas” (BUTTERICK, 2023).

Em uma entrevista ao site de notícias Forbes (2022), o fundador e CEO da Midjourney, David Holz, responde algumas perguntas sobre consentimento dos artistas em relação à arte usada para treinar sua IA. Em resposta, Holz afirma não ser possível identificar o proprietário de cada imagem utilizada, já que o *dataset* utilizado para treinar a IA contém mais de 100 milhões de itens presentes na internet. Ao ser perguntado sobre a possibilidade de artistas escolherem que sua arte não seja utilizada para o treinamento, o CEO afirma ser difícil autenticar se a pessoa que diz ser o artista é realmente dono da arte além de conseguir encontrar a arte no *dataset*. Holz afirma que a Midjourney não foi feita para substituir profissionais em seu trabalho, mas sim dar a possibilidade de qualquer pessoa expressar sua criatividade (FORBES, 2022).

Apesar da visão humanista do CEO, é inegável os riscos que a IA generativa apresenta a artistas seja como competição “injusta” ou infringimento de direitos autorais, além da facilidade com que a mesma facilita outros problemas potencialmente graves como fabricação de mídia sintética com agendas pessoais ou políticas. Holz na entrevista diz que “a tecnologia em si não é o problema. É como água. A água pode ser perigosa, você pode se

afogar nela. Mas também é essencial" (SALKOWITZ, 2022), referenciando a questão da tecnologia ser uma "faca de dois gumes", com seu perigo sendo como é usada.

Esta ausência de responsabilidade apresenta diversos riscos. Primeiramente, pelos avanços na tecnologia estarem mais rápidos do que as formas de as controlar e segundo que suas possibilidades excedem as expectativas dos próprios desenvolvedores que as criam.

Um exemplo dessa utilização que transcende as expectativas dos desenvolvedores é a utilização do modelo transformativo de texto ChatGPT. O ChatGPT já se encontra como um conhecimento comum na cultura popular brasileira e mundial. A IA que consegue criar linhas de texto com precisão semelhante a diálogos humanos traz consigo discussões polêmicas sobre plágio. Em seu artigo sobre a ética da IA generativa, King, McMillan e Zohny (2023) afirmam que a utilização do ChatGPT no âmbito acadêmico é perigosa pois, apesar do modelo criar texto que aparenta conciso e bem desenvolvido, o sistema muitas vezes tende a plagiar textos disponíveis online, não tem a capacidade de citar referências de forma adequada e suas respostas dependem completamente das perguntas feitas.

Apesar da facilidade que o ChatGPT traz para "embelezar" o texto, King, McMillan e Zohny (2023) ressaltam a importância dos momentos de esforço que um autor passa ao tentar expressar suas ideias em texto pode acarretar em epifanias e como pular este passo pode levar a potenciais epifanias perdidas.

O plágio e a morte da criatividade não são os únicos problemas apresentados pela presença do ChatGPT na sociedade. Em 2018, o professor de direito Jonathan Thurley foi informado pelo jurista americano Eugene Volokh que seu nome apareceu em uma lista de juristas fornecidas pelo ChatGPT. Volokh estava pesquisando a capacidade do modelo transformativo de fornecer dados incorretos e reconheceu o nome de seu colega (GARATTONI, 2023).

> "Quando pedi as fontes da informação, o ChatGPT fez pior ainda. Além de citar o tal artigo do Washington Post, que não existiu ('GWU Law professor accused of sexual harassment by former student', dez/2018), também mencionou reportagens da emissora NBC News e do The National Law Journal, bem como uma suposta defesa que teria sido publicada no blog do próprio Turley. Tudo completamente fictício" (GARATTONI, 2023).

Esta falsa notícia, que supostamente haveria ocorrido na década de 1990, apresentou riscos graves à reputação do professor e mostrou os verdadeiros perigos de uma ferramenta antes vista como "inofensiva". O fenômeno que levou a esse acontecimento é descrito por Garattoni (2023) como um caso de "alucinação", onde modelos de linguagem transformativos como o ChatGPT e outros, que picotam e recombinam palavras para gerar

suas frases, acabam por criar combinações que não possuem nenhuma relação com fatos reais, mas que parecem reais o suficiente para enganar possíveis leitores.

Esta não é a primeira vez que a IA produz informações falsas que difamam um sujeito, com Brian Hood, prefeito de Hepburn Shire, Austrália, tendo sido vítima de notícias falsas sobre ele. Em seu caso, Hood foi descrito como condenado por um escândalo de propina no Reserve Bank of Australia, nos anos 2000 (GARATTONI, 2023).

Entretanto, os problemas causados pelo uso de IA generativa não se resumem exclusivamente a problemas da máquina. Em 2023 um funcionário de um escritório de advocacia em Nova York utilizou o ChatGPT para encontrar casos legais que seriam favoráveis ao seu cliente, a fim de apresentar casos semelhantes com veredictos que beneficiassem as vítimas. Ao pesquisar os casos, foi confirmado que os mesmos nem sequer existiam (EXAME, 2023) e, ao ser confrontado sobre o caso, o advogado confirmou ter utilizado a IA e que havia questionado a mesma sobre a veracidade de um dos casos, para qual a ferramenta respondeu que “sim, esse é um caso real” (MARUF, 2023).

Seguindo a repercussão causada por casos como este, professores de ética legal ao redor do mundo ressaltam os problemas da utilização de inteligências artificiais no meio jurídico e a necessidade da verificação do conteúdo gerado por elas (WEISER, 2023).

Com a evolução da síntese de informação se apresenta um grande desafio para a identificação do uso de IA para geração de mídia sintética. Angelis et al. (2023) apontam os perigos que a tecnologia apresenta: mesmo com travas e filtros criados pelos seus programadores, modelos de linguagem transformativos de grande escala podem ser manipulados através de texto para exercerem funções as quais foram programadas para explicitamente não fazerem.

Compactuando com a visão de que as IAs generativas apresentam potenciais perigos, Geoffrey Hinton (2023), ex-vice-presidente de engenharia do Google e renomado no meio da computação, aponta que questões como o impacto no desemprego devido à substituição de humanos por máquinas e a geração de notícias falsas são pontos que já estão causando discussões. Essas discussões o fizeram mudar sua previsão anterior de que a inteligência artificial poderia superar a inteligência humana em trinta a cinquenta anos para cinco a vinte anos, baseado na velocidade de evolução da área. Hinton também menciona a possibilidade de um perigo bélico estar associado ao progresso das IAs, como soldados robóticos inteligentes. Por fim, ele orienta que sejam direcionados esforços para conter os perigos que acompanham a evolução das IAs, como, por exemplo, desenvolver uma forma de

diferenciar imagens falsas das reais com base em alguma forma de marcação e estabelecer um plano de contingência para o caso de surgir alguma ameaça imediata (UNISINOS, 2023).

Com o advento da automatização avançada dos modelos de inteligência artificial, também houveram polêmicas relacionadas a moderação de conteúdo criado com ferramentas automatizadas feita para empresas como Facebook, Tiktok e ChatGPT, estas que utilizam modelos de inteligência artificial em seus serviços. Esta moderação, exercida por meio de outsourcing de trabalhadores com salários tão baixos quanto US$ 1,50 por hora, levou os mesmos a formarem o “Sindicato Africano dos Moderadores de Conteúdo" para lutar por seus direitos (PERRIGO, 2023).

A moderação feita por meio de mão de obra considerada barata se tornou alvo dos holofotes quando a informação que a empresa responsável pelo ChatGPT, OpenAI, utilizou em 2021 os serviços de uma empresa situada no Kenya, África, chamada Sama, para rotular os textos enviados pela OpenAI para que fossem utilizados para integrar um sistema de filtro ao ChatGPT contra conteúdo tóxico. O trabalho, o qual requer que os trabalhadores revisem e façam a categorização dos prompts de texto não-filtrados, proporciona condições ruins de trabalho pois além de oferecer salários variando de US$ 1,32 até US$ 2,00, os colaboradores foram expostos a textos e imagens representando atos obscenos e perturbadores, como por exemplo descrições de atos libidinosos com animais e menores de idade. A exposição a tais informações causou impactos devastadores na saúde mental dos funcionários, como em casos onde alguns moderadores desenvolveram transtorno de estresse pós-traumático. Devido a isto, a Sama cancelou o contrato 8 meses antes do que o estipulado pois o acordo inicial não mencionava o envio de conteúdo ilegal, ainda mais considerando que parte destes conteúdos foram classificados, em termos usados pela OpenAI, como “C4”, que constitui conteúdo relacionado a abuso sexual infantil e também “C3’ e “V3”, denotando cenas como estupro e morte, respectivamente (PERRIGO, 2023).

A quantidade de horas e o volume de conteúdo criado diariamente além da falta de moralidade presente nos sistemas automáticos leva a preocupante necessidade da conscientização de pesquisadores e desenvolvedores na área da inteligência artificial. Responsabilidade, tanto legal quanto moral e desenvolvimento não só de avanços na criação de mídia sintética, mas também nas formas de explicitamente identificar estas criações, coibir a criação de conteúdo ofensivo, falso e difamatório, e melhor controlar os comportamentos dos agentes inteligentes que cada vez mais moldam falsas realidades é um dever dos criadores.

Assim como Robert Oppenheimer, diretor do Projeto Manhattan, depois arrependeu-se de sua criação, a bomba atômica, uma ferramenta que tirou milhões de vidas (RODRÍGUEZ, 2023), criadores e pesquisadores de IA devem ter a noção que suas criações terão marcas permanentes na sociedade e cultura humana, podendo até mesmo moldar os próximos passos do nosso desenvolvimento como espécie.

5. CONCLUSÕES E TRABALHOS FUTUROS

O presente estudo teve como objetivo principal realizar uma abrangente discussão e levantamento de dados sobre a história, evolução e estado atual da inteligência artificial, com ênfase na criação de conteúdo por meio de sistemas inteligentes.

Com base nos resultados obtidos, pode-se concluir que, por se tratar de um campo relativamente novo, a criação de mídias por inteligências artificiais ainda está em processo de amadurecimento. Questões éticas e legais emergem nesse contexto, como a necessidade de adaptação ou criação de leis relacionadas ao plágio envolvendo imagens não autorizadas ou a geração de textos plagiados e/ou potencialmente falsos por modelos de geração de texto.

Diante da falta de regulamentação atual, torna-se evidente a urgência de estabelecer leis e regulamentações que abordem o crescente uso de inteligências artificiais na criação de mídias sintéticas, a fim de prevenir controvérsias como plágio, ataques pessoais e políticos, disseminação em massa de conteúdo falso e desinformação.

É plausível supor que, à medida que novas leis e regulamentações sejam promulgadas nessa área, possam surgir cargos específicos relacionados a inteligências artificiais e mídias sintéticas, com foco na ética e conformidade com as normas estabelecidas, semelhante ao que ocorreu com a implementação da Lei Geral de Proteção de Dados. A rápida adaptação e atenção a essas demandas são essenciais para garantir um uso responsável e benéfico das tecnologias envolvidas.

5.1. TRABALHOS FUTUROS

Esse trabalho sugere, por fim, alguns direcionamentos que podem agregar enriquecimento à pesquisa aqui elaborada. Seguem eles:

- Estudo e desenvolvimento de uma inteligência artificial que possa identificar se uma mídia foi criada por um humano ou por uma inteligência artificial;
- Estudo específico sobre a evolução da inteligência artificial na área da educação e seus efeitos no aprendizado dos estudantes;

- Discussão sobre o impacto do uso de inteligências artificiais na capacidade criativa humana.

REFERÊNCIAS

AMAZON. **O que é uma rede neural?** Disponível em: <https://aws.amazon.com/pt/what-is/neural-network/>. Acesso em: 08 dez. 2022

ARIMETRICS (Espanha) (org.). **What is Midjourney**. 2022. Disponível em: https://www.arimetrics.com/en/digital-glossary/midjourney. Acesso em: 28 maio 2023.

ARNTZ, Melanie; GREGORY, Terry; ZIERAHN, Ulrich. Revisiting the risk of automation. **Economics Letters**, [S.L.], v. 159, n. 1, p. 157-160, out. 2017. Elsevier BV. http://dx.doi.org/10.1016/j.econlet.2017.07.001.

BERKELEY SCHOOL OF INFORMATION. **What Is Machine Learning (ML)?** Disponível em: <https://ischoolonline.berkeley.edu/blog/what-is-machine-learning/>. Acesso em: 08 dez. 2022.

BERNDTSSON, Mikael; MELLIN, Jonas. ECA Rules. **Encyclopedia Of Database Systems**, [S.L.], p. 959-960, 2009. Springer US. http://dx.doi.org/10.1007/978-0-387-39940-9_504

BRASIL. Câmara dos Deputados. Projeto de Lei nº 21/20, de 04 de fevereiro de 2020. Estabelece fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da inteligência artificial no Brasil; e dá outras providências. Brasília: Câmara dos Deputados, 2020. Disponível em: <https://www.camara.leg.br/propostas-legislativas/2236340>. Acesso em: 08 dez. 2022.

BRITANNICA, The Editors Of Encyclopaedia. **Bayesian analysis**. 2023. Disponível em: https://www.britannica.com/science/Bayesian-analysis. Acesso em: 29 maio 2023

BUCH, Varun H; AHMED, Irfan; MARUTHAPPU, Mahiben. Artificial intelligence in medicine: current trends and future possibilities. **British Journal Of General Practice**, [S.L.], v. 68, n. 668, p. 143-144, 22 fev. 2018. Royal College of General Practitioners. http://dx.doi.org/10.3399/bjgp18x695213.

BURCHARD, Hans von Der. **Belgian socialist party circulates 'deep fake' Donald Trump video**. 2018. Disponível em: https://www.politico.eu/article/spa-donald-trump-belgium-paris-climate-agreement-belgian-socialist-party-circulates-deep-fake-trump-video/. Acesso em: 29 maio 2023.

CARVALHO, Rodrigo César Picon de. **Deepfake, inteligência artificial e Direito Penal Eleitoral**. 2022. Disponível em: https://canalcienciascriminais.com.br/deepfake-inteligencia-artificial-e-direito-penal-eleitoral/. Acesso em: 17 maio 2023.

CHEN, Xieling ; ZOU, Di ; XIE, Haoran et al. / **Two decades of artificial intelligence in education: Contributors, collaborations, research Topics, challenges, and future directions**. In: Educational Technology and Society. 2022 ; Vol. 25, No. 1. pp. 28-47.

EDA KAVLAKOGLU. **AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the Difference?** 2020. Disponível em: <https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks>. Acesso em: 08 dez. 2022.

European Commission. **LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS**. Disponível em: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>. Acesso em: 08 dez. 2022.

EXAME. **ChatGPT é usado por advogado em processo, mas fornece informações falsas**. 2023. Disponível em: https://exame.com/future-of-money/chatgpt-usado-advogado-processo-fornece-informacoes-falsas/. Acesso em: 30 maio 2023.

FIA BUSINESS SCHOOL. **Machine Learning: como funciona, benefícios, tipos e exemplos.** Disponível em: <https://fia.com.br/blog/machine-learning/>. Acesso em: 08 dez. 2022.

GARATTONI, Bruno. **ChatGPT acusa professor de assédio sexual – e usa como prova reportagens que não existiram Leia mais em: https://super.abril.com.br/coluna/bruno-garattoni/chatgpt-acusa-professor-de-assedio-sexual-e-cita-reportagens-que-nao-existiram**. 2023. Disponível em: https://super.abril.com.br/coluna/bruno-garattoni/chatgpt-acusa-professor-de-assedio-sexual-e-cita-reportagens-que-nao-existiram. Acesso em: 29 maio 2023.

GENERATIVEAI. Generative AI: first, what is generative ai?. First, What is Generative AI?. Disponível em: https://generativeai.net/. Acesso em: 18 jun. 2023.

HOLMES, Wayne *et al*. **Artificial Intelligence In Education**: promises and implications for teaching and learning. Boston: Center For Curriculum Redesign, 2019. 39 p.

IBM CLOUD EDUCATION. **O que é machine learning?** 2020. Disponível em:<https://www.ibm.com/br-pt/cloud/learn/machine-learning>. Acesso em: 08 dez. 2022.

IBM CLOUD EDUCATION. **O que é o deep learning?** 2020. Disponível em: <https://www.ibm.com/br-pt/cloud/learn/deep-learning>. Acesso em: 08 dez. 2022.

IBM CLOUD EDUCATION. **O que são Redes Neurais?** 2020. Disponível em: <https://www.ibm.com/br-pt/cloud/learn/deep-learning>. Acesso em: 08 dez. 2022.

JOHNSON, Dave *et al*. **What are deepfakes?**: how fake ai-powered media can warp our perception of reality. How fake AI-powered media can warp our perception of reality. 2023. Disponível em: https://www.businessinsider.com/guides/tech/what-is-deepfake. Acesso em: 29 maio 2023

LEARNING, Mit Open. **Tackling the misinformation epidemic with "In Event of Moon Disaster"**: new website from the mit center for advanced virtuality rewrites an important moment in history to educate the public on the dangers of deepfakes. New website from the

MIT Center for Advanced Virtuality rewrites an important moment in history to educate the public on the dangers of deepfakes. 2020. Disponível em: https://news.mit.edu/2020/mit-tackles-misinformation-in-event-of-moon-disaster-0720. Acesso em: 29 maio 2023.

MARUF, Ramishah. **Lawyer apologizes for fake court citations from ChatGPT**. 2023. Disponível em: https://edition.cnn.com/2023/05/27/business/chat-gpt-avianca-mata-lawyers/index.html. Acesso em: 30 maio 2023.

MCCARTHY, John. **What is artificial inteligence?** Stanford: Stanford University, 2007. Disponível em: <http://jmc.stanford.edu/articles/whatisai/whatisai.pdf>. Acesso em: 08 dez. 2022.

MCKINSEY & COMPANY. **Insights to impact**: creating and sustaining data-driven commercial growth. Creating and sustaining data-driven commercial growth. 2022. Disponível em: <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/insights-to-impact-creating-and-sustaining-data-driven-commercial-growth>. Acesso em: 08 dez. 2022.

MIDJOURNEY. **Direct Messages**. 2022. Disponível em: https://openai.com/blog/chatgpt. Acesso em: 28 maio 2023.

MOUSAVINASAB, Elham; ZARIFSANAIEY, Nahid; KALHORI, Sharareh R. Niakan; RAKHSHAN, Mahnaz; KEIKHA, Leila; SAEEDI, Marjan Ghazi. Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods. **Interactive Learning Environments**, [S.L.], v. 29, n. 1, p. 142-163, 18 dez. 2018. Informa UK Limited. http://dx.doi.org/10.1080/10494820.2018.1558257.

OPENAI. **DALL·E.** Disponível em: https://openai.com/dall-e/. Acesso em: 12 abr. 2023.

OPENAI. **Introducing ChatGPT**. 2022. Disponível em: https://openai.com/blog/chatgpt. Acesso em: 28 maio 2023.

OPENAI. **Language Models are Few-Shot Learners**. Baltimore, 2020. Disponível em: <https://arxiv.org/pdf/2005.14165.pdf>. Acesso em: 08 dez. 2022.

PERRIGO, Billy. **150 African Workers for ChatGPT, TikTok and Facebook Vote to Unionize at Landmark Nairobi Meeting**. 2023. Disponível em: https://time.com/6275995/chatgpt-facebook-african-workers-union/. Acesso em: 30 maio 2023.

PERRIGO, Billy. **Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic**. 2023. Disponível em: https://time.com/6247678/openai-chatgpt-kenya-workers/. Acesso em: 30 maio 2023.

PRISMA LABS. **Privacy Policy.** 2022. Disponível em: https://tos.lensa-ai.com/privacy. Acesso em: 16 maio 2023.

RAMESH, An; KAMBHAMPATI, C; MONSON, Jrt; DREW, Pj. Artificial intelligence in medicine. **Annals Of The Royal College Of Surgeons Of England**, [S.L.], v. 86, n. 5, p. 334-338, 1 set. 2004. Royal College of Surgeons of England. http://dx.doi.org/10.1308/147870804290.

RODRÍGUEZ, Héctor. **Robert Oppenheimer, o pai arrependido da bomba atómica**. 2023. Disponível em: https://nationalgeographic.pt/historia/actualidade/3782-robert-oppenheimer-o-pai-arrependido-da-bomba-atomica. Acesso em: 31 maio 2023.

RUSSELL, Stuart; NORVIG, Peter. **Inteligência Artificial**: tradução da terceira edição. 3. ed. Rio de Janeiro: Elsevier, 2013.

S. Srivastava, A. Bisht and N. Narayan, "Safety and security in smart cities using artificial intelligence — A review," 2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence, Noida, India, 2017, pp. 130-133, doi: 10.1109/CONFLUENCE.2017.7943136.

SALIAN, Isha. **SuperVize Me: What's the Difference Between Supervised, Unsupervised, Semi-Supervised and Reinforcement Learning?** 2018. Disponível em: <https://blogs.nvidia.com/blog/2018/08/02/supervised-unsupervised-learning/>. Acesso em: 08 dez. 2022.

SALKOWITZ, Rob. **Midjourney Founder David Holz On The Impact Of AI On Art, Imagination And The Creative Economy**. 2022. Disponível em: https://www.forbes.com/sites/robsalkowitz/2022/09/16/midjourney-founder-david-holz-on-the-impact-of-ai-on-art-imagination-and-the-creative-economy/?sh=9876bf22d2b8. Acesso em: 29 maio 2023.

SOMERS, Meredith. **Deepfakes, explained**. 2020. Disponível em: https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained. Acesso em: 17 maio 2023.

SOTTILE, Zoe. **What to know about Lensa, the AI portrait app all over social media**. 2022. Disponível em: https://edition.cnn.com/style/article/lensa-ai-app-art-explainer-trnd/index.html. Acesso em: 29 maio 2023.

STUPP, Catherine. **Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case**: scams using artificial intelligence are a new challenge for companies. Scams using artificial intelligence are a new challenge for companies. 2019. Disponível em: https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402. Acesso em: 17 maio 2023.

THOMPSON, Anne. **Sundance 2020**: how vfx pulled 'welcome to chechnya' out of the shadows. How VFX Pulled 'Welcome to Chechnya' Out of the Shadows. 2020. Disponível em: https://www.indiewire.com/video/sundance-2020-vfx-hbo-welcome-to-chechnya-documentary-david-france-1202204703/. Acesso em: 29 maio 2023.

TOEWS, Rob. **Deepfakes Are Going To Wreak Havoc On Society. We Are Not Prepared**. 2020. Disponível em: https://www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/?sh=9f7994274940. Acesso em: 29 maio 2023.

TURING, A. M.. I.—COMPUTING MACHINERY AND INTELLIGENCE. **Mind**, [S.L.], v. , n. 236, p. 433-460, 1 out. 1950. Oxford University Press (OUP). http://dx.doi.org/10.1093/mind/lix.236.433.

UNISINOS, Instituto Humanitas. **“Se existe alguma maneira de controlar a inteligência artificial, devemos descobri-la antes que seja tarde demais”. Entrevista com Geoffrey Hinton**. 2023. Disponível em: https://www.ihu.unisinos.br/628477-se-existe-alguma-maneira-de-controlar-a-inteligencia-artificial-devemos-descobri-la-antes-que-seja-tarde-demais-entrevista-com-geoffrey-hinton. Acesso em: 30 maio 2023.

WEISER, Benjamin. **Here’s What Happens When Your Lawyer Uses ChatGPT**. 2023. Disponível em: https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html. Acesso em: 30 maio 2023.

WITNESS.ORG. **INTRODUCING THE WITNESS MEDIA LAB**. 2015. Disponível em: https://www.witness.org/introducing-witness-media-lab/. Acesso em: 29 maio 2023.

ZOHNY, Hazem; MCMILLAN, John; KING, Mike. Ethics of generative AI. **Journal Of Medical Ethics**, [S.L.], v. 49, n. 2, p. 79-80, 24 jan. 2023. BMJ. http://dx.doi.org/10.1136/jme-2023-108909.